UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Maternal breastfeeding in Portugal: factors associated with the duration and premature cessation

Sílvia Cristina Martins Mendonça

TRABALHO DE PROJETO
MESTRADO EM BIOESTATÍSTICA

2013

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Maternal breastfeeding in Portugal: factors associated with the duration and premature cessation

Sílvia Cristina Martins Mendonça

TRABALHO DE PROJETO
MESTRADO EM BIOESTATÍSTICA

Orientadora:
Professora Doutora Marília Cristina de Sousa Antunes

2013

Resumo

Embora seja conhecido que o aleitamento materno é a melhor forma de alimentar o recém-nascido, as taxas de amamentação ficam frequentemente abaixo das recomendadas pelas autoridades de saúde pública.

A amamentação é um fenómeno complexo, influenciado por características demográficas e psicológicas maternas bem como constrangimentos biológicos e apoio social. Aumentar as taxas de amamentação torna-se, portanto, um processo difícil.

Esta tese contribui para a compreensão da problemática da amamentação em Portugal através da análise de dados observacionais sobre práticas da amamentação e seus determinantes.

O modelo de regressão de Cox foi usado para identificar determinantes demográficos, biológicos, sociais e psicológicos da duração da amamentação. A modelação teve em consideração as relações causais e temporais entre os vários níveis de determinantes. Este modelo identificou factores demográficos já conhecidos, como a idade escolaridade maternas, mas também factores modificáveis associados aos cuidados de saúde e ao conhecimento materno, tais como a introdução precoce de leite artificial e chupeta e o conhecimento de técnicas de amamentação.

O modelo de regressão logística foi usado para estudar o efeito do aconselhamento pelos profissionais de saúde no sucesso na amamentação, definido como o atingimento da duração planeada da amamentação. Em particular, as variáveis de exposição referem-se ao aconselhamento sobre benefícios, técnicas e dificuldade da amamentação. Este modelo estimou um efeito positivo, com significância marginal, do aconselhamento sobre benefícios no período pré-natal. As principais limitações desta análise prendem-se com o facto de os dados serem observacionais e de as questões colocadas serem pouco específicas.

Palavras Chave: Regressão de Cox, Regressão logística, Aleitamento materno, Duração do aleitamento, Atingimento da meta

Abstract

Despite an extensive body of evidence that breastfeeding (BF) is the most adequate way to feed the newborn, BF rates frequently fall below the public health recommendations. It is a complex phenomenon influenced by maternal demographic and psychological characteristics as well as biomedical constraints and social support. Improving BF rates is therefore a difficult process.

This thesis contributes to the understanding of BF in Portugal through the analysis of observational data on BF practices and determinants.

A Cox regression model was used to identify demographic, biological, social and psychological determinants of BF duration. Model building took into account the causal and temporal relations among the different levels of determinants. This model identified well known demographic factors, such as maternal age and education, but also modifiable factors associated with the health care system and maternal knowledge, such as early introduction of formula and pacifier and the knowledge of BF techniques.

Logistic regression was used to study the effect of counselling by health professionals on BF success, defined as the achievement of the intended BF duration. In particular, the exposure variables refer to counselling about BF benefits, techniques and difficulties. This model estimated a positive effect, with borderline significance, of counselling about BF benefits in the antenatal period. The main limitations of this analysis are the observational nature of the data and the questions applied being generic.

Keywords: Cox regression, Logistic regression, Maternal breastfeeding, Breastfeeding duration, Goal achievement

Resumo Alargado

As recomendações de saúde pública sobre o aleitamento materno são suportadas por evidência científica de benefícios tanto para a mãe como para o bebé. No entanto, as taxas de amamentação (AM) ficam frequentemente abaixo das recomendadas.

A amamentação é um fenómeno complexo, influenciado por características demográficas e psicológicas maternas bem como constrangimentos biológicos e apoio social. Aumentar as taxas de amamentação torna-se, portanto, um processo difícil.

Esta tese contribui para a compreensão da problemática da amamentação em Portugal através da análise de dados observacionais sobre práticas da amamentação e seus determinantes. Foram inicialmente incluídos 494 pares mãe-bebé selecionados aleatoriamente de entre os nascimentos ocorridos entre dezembro de 2010 e janeiro de 2011. Os dados foram recolhidos por entrevista telefónica aproximadamente aos três, seis e dezoito meses de idade do bebé.

Em particular, esta tese propôs-se a: 1) identificar fatores associados à duração do aleitamento através de um modelo de regressão de Cox e 2) estudar o efeito do aconselhamento dado pelos profissionais de saúde no atingimento da meta da amamentação através de uma regressão logística.

O modelo de Cox exprime a função de risco como um produto de uma função do tempo, a função de risco basal, e uma função das covariáveis. A medida de efeito é a razão dos riscos e corresponde à razão entre as funções de risco de indivíduos com valores diferentes de covariáveis.

A variável resposta do modelo de análise de sobrevivência foi a *duração de qualquer aleitamento* (qqAM). De entre as 473 mães que iniciaram o aleitamento, ocorreu censura em 10.5% dos casos por perda de seguimento e noutros 8.8% que estavam a amamentar aos dezoito meses. As estimativas de Kaplan-Meier para o tempo mediano de sobrevivência foram de 2 e 5 meses para a amamentação em exclusivo e qqAM, respetivamente.

As variáveis independentes associadas à duração de qqAM foram agrupadas como fatores demográficos, biológicos, sociais e psicológicos. O modelo de Cox foi construído de forma hierárquica, em que as variáveis pertencentes a cada bloco foram consideradas para entrada no modelo nesta ordem específica. Dentro de cada bloco, as variáveis foram selecionadas por um método *stepwise*, tendo-se usado o teste de razão de verosimilhanças para comparação de modelos encaixados. A hierarquização dos blocos tenta refletir as relações causais e temporais entre os vários grupos de determinantes e, desta forma, assegurar o ajuste por variáveis de confundimento.

Foram usados os resíduos de Shoenfeld para avaliar o pressuposto da proporcionalidade dos riscos. A forma funcional das covariáveis contínuas, *idade materna* e *BSES* foi averiguada usando os resíduos de martingale. Foi investigada a transformação da variável *idade materna* através de uma função P-*spline* que, no entanto, se revelou desnecessária. Os resíduos de *deviance* foram usados para a identificação de observações mal ajustadas pelo modelo e os resíduos de Score para a identificação de observações influentes.

De entre os fatores demográficos, o modelo final inclui a idade materna, escolaridade e país de origem, todos com um nível de significância do teste de Wald abaixo dos 10%. Entre as variáveis biológicas destacam-se as diferenças encontradas entre o tipo de parto, apesar de não serem significativas no modelo final, e o efeito negativo das complicações da amamentação. A ocorrência de mastite está associada a um aumento do risco de cessação de quase 70%.

A introdução precoce de chupeta, antes do primeiro mês de vida, está associada a um aumento do risco de 39% e revela significância estatística no modelo final. No entanto, a sua relação com o fim da AM não é consensual na literatura.

Finalmente, vale a pensa destacar que duas variáveis psicológicas associadas ao conhecimento materno, o número de técnicas de AM reportadas e o reconhecimento do contributo da AM para a relação mãe-bebé, permanecem no modelo com significância estatística, ambas associadas a uma diminuição no risco de 20%. O modelo final explica 36.3% da variabilidade total na variável resposta, entre os quais 21% correspondem às variáveis psicológicas incluídas. A probabilidade de concordância é de 0.735 e portanto, muito boa.

O modelo logístico é um modelo linear generalizado que relaciona o valor esperado de uma variável binária com o preditor linear através da transformação *logit*. A medida de efeito é a razão das chances e corresponde à chance de um determinado evento ocorrer, dada a exposição a um fator de risco, comparada com a chance na ausência de exposição. A variável resposta *atingimento da meta* é uma variável binária que define o sucesso do aleitamento através do atingimento da meta para a duração da AM estabelecida individualmente, como alternativa a metas temporais que são geralmente usadas.

Foram consideradas 452 mães que especificaram uma meta quantitativa para a duração da AM. De entre estas, a maioria referiu seis (23.2%) e 12 meses (45.4%) como duração planeada. Devido à censura apenas é possível determinar a variável resposta para 411 casos, dos quais aproximadamente um terço atingiram a meta. No entanto, é provável que a proporção real de sucessos seja superior à encontrada, devido à perda diferencial de indivíduos que amamentaram mais tempo.

As variáveis de interesse deste modelo correspondem a duas questões onde foi perguntado se alguma vez tinham sido aconselhadas sobre benefícios da AM e sobre técnicas ou dificuldades da AM. As variáveis foram categorizadas como *Nunca*, *Período pré-natal* e *Período pós-natal*. A maioria das mães recorda ter aconselhada pelo menos uma vez (77% sobre benefícios e 71% sobre técnicas/dificuldades). Existem diferenças significativas entre os dois grupos em variáveis como a idade, escolaridade e experiência prévia de AM.

Na modelação, as variáveis *idade materna*, *educação*, *experiência prévia de AM* e *país de origem* foram consideradas variáveis de confundimento *a priori* e incluídas independentemente da significância estatística. A entrada de outras variáveis biológicas dependeu da associação com a variável resposta e existência de efeito de confundimento, medida como a variação na medida de efeito.

A observação dos resíduos de *deviance* estandardizados sugeriu a existência de uma relação não linear entre a *idade materna* e o *logit*. Esta variável foi transformada com recurso a polinómios de graus dois e três e através de um *spline* cúbico. Optou-se por corrigir a não linearidade com um *spline* cúbico com dois nós internos. O aumento da idade materna leva a um aumento no valor ajustado para a probabilidade de sucesso até cerca dos 30 anos, após os quais não há alteração dos valores ajustados com o aumento da idade.

O teste de Hosmer-Lemershow revela que o modelo tem um bom ajuste. A área debaixo da curva ROC é de quase 70% pelo que a capacidade discriminatória está muito próxima da aceitável.

De entre as variáveis de interesse, apenas o aconselhamento sobre benefícios exibiu uma associação com o sucesso na AM com significância marginal (p=0.1165). O aconselhamento sobre benefícios no período pré-natal está associado a um aumento de 72% na *chance* do sucesso, comparativamente à ausência desse aconselhamento.

A idade materna e o país de origem são significativos no modelo final. A existência de complicações como bloqueio de ductos, mastite ou fissuras está associada à redução de 45% na *chance* de sucesso. A experiência prévia de AM, apesar de ter revelado associação significativa com o sucesso quando analisada de forma independente, não permanece significativa no modelo final.

A importância do aconselhamento e apoio por profissionais de saúde, tanto no período pré-natal como no pós-natal, já foi demonstrada em estudos intervencionais. As principais limitações desta análise prendem-se com o facto de os dados serem observacionais e de as questões colocadas serem pouco específicas.

Acknowledgements

Writing these acknowledgements marks the end of two years of dedication to this masters degree. This work would not have been possible without the help and encouragement of many people.

First of all, I wish to thank my supervisor Prof. Marília Antunes, for her patience, guidance and hard work during the last year.

I also want to thank IMP for the opportunity to work on a real epidemiology dataset. In particular, I wish to thank Violeta Alarcão and Rui Simões for the suggestions and help in understanding the dataset and the theme.

I wish to acknowledge the good spirit and friendship among my classmates and all the excellent teachers I had in these two years, with a special mention to Prof. Lisete Sousa, for her availability and hard work.

Finally, I want to thank my parents for supporting my decision to go back to study and to Tiago for all the support and understanding during the last two years.

Sílvia Mendonça, July 2013

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Public health recommendations on maternal breastfeeding (BF) are supported by an extensive body of evidence that includes benefits for both mother and child (Village, 2012).

Breast milk is easier to digest than formula milk and has the right combination of nutrients to feed the newborn. It also contains components of the maternal immune system that are able to confer immune protection and stimulate the development of the baby's immune system (Eidelman *et al.*, 2012; Hanson, 2000). In particular it is associated with a decreased risk of respiratory and gastrointestinal infections, as well as inflammatory and allergic diseases. On the mothers' side, BF contributes to a more rapid return to their pre-pregnancy weight and increased child spacing through lactational amenorrhoea. It has also been linked with a decreased risk of breast cancer and this effect is proportional to the cumulative duration (Eidelman *et al.*, 2012).

Based on this evidence, the World Health Organization (WHO) recommends exclusive BF up to six months of age. However for most developed countries BF rates, both exclusive and complemented, are far below this goal. In Portugal, the National Health Plan for 2004-2010 had set the target of 50% exclusive BF at least until three months of age. Unpublished results from this study as well as other studies (Orfão, 2012) show that while BF initiation rates are high (above 98%), exclusive BF at three months does not exceed 40%.

Improving BF rates is a difficult process due to its multi factorial nature. Maternal demographic and psychological determinants, biomedical characteristics, health care support and cultural acceptance and support can all influence the initiation and duration of BF (Hector *et al.*, 2005; Thulier & Mercer, 2009; Yngve & Sjöström, 2001). The classification of BF determinants as demographic, biological, social and psychological is used in this thesis (Thulier & Mercer, 2009). Demographic determinants include maternal age, level of education, marital status, race and socioeconomic status. Biological variables include the delivery method, parity, the existence of infant health problems, BF difficulties and complications. Social variables relate to the working status of the mother, support given by family and the health care system, including the pre-natal period, maternity care and primary care in the follow-up period. They include maternity ward practices such as early initiation and avoidance of pacifiers and tetins. Psychological determinants relate to the mother's intention to breastfeed, her knowledge, skills and confidence in her ability to breastfeed.

Studies aiming to identify BF determinants apply mainly two statistical models: logistic regression and survival analysis. Both models have advantages and disadvantages and the choice between them depends on the data available and, essentially, on the study purpose.

Logistic regression models the probability of a binary outcome, usually classified as success or failure. This analysis requires the outcome to be known for all individuals. It would be the adequate choice to model BF data available for only a specific point in time (cross-sectional data). However, if information is available for the full duration of BF for all or most of the individuals, it would be necessary to consider a specific point in time as the cutpoint to distinguish failure from success. The cutpoint most frequently used in the literature for any BF is six months; other cutpoints include three, four and twelve months of baby's age (Al-Sahab *et al.*, 2010; Caldeira *et al.*, 2000; Forster *et al.*, 2006; Rasenack *et al.*, 2012). Using this criteria, BF beyond the defined point e.g., six months, would be the success.

Survival analysis models time to event data. It can accommodate censured data which means, when partial information about the duration is available, it can be used for modelling. Survival analysis models for BF duration are less common in the literature than logistic regression and the ones found use Cox regression (Bai *et al.*, 2013; Scott *et al.*, 2006; Skafida, 2012).

1.2 Objectives

The main objectives of this thesis are:

1. to identify factors associated with BF duration. This is accomplished through the use of a Cox regression model.

2. to investigate the effect of specific variables related to the transmission of knowledge from health care professionals on the achievement of the individual planned BF duration. This is achieved by a logistic regression model.

1.3 Overview

The overview of this document is as follows.

Chapter 2 introduces the data and the project where it originates from. It also provides details about how the response variables were constructed.

Chapters 3 and 5 explain the basic theory for model formulation, estimation and diagnostics for Cox and logistic regression, respectively.

Chapter 4 describes the application of a Cox model to identify factors associated with the duration of any BF.

Chapter 6 describes the construction of a logistic regression model to study the effect of BF counselling on BF success.

Finally, chapter 7 summarizes the main conclusions from this work.

Chapter 2

Database

2.1 Project

The data used in this thesis was collected as part of a project carried out at the Institute of Preventive Medicine (IMP-FMUL) that aimed to characterize BF in the portuguese context, determine its prevalence and determinants. The study was observational and longitudinal with sampling on a national level. Data was collected in three telephone interviews at approximately three, six and eighteen months of babies age. The three month questionnaire is shown in Appendix D. Only the first two contacts were initially planned. The last contact was made in order to complete as much as possible the information about the total duration of BF, without which the analysis carried out in this thesis would not have been possible.

Children included were born between December 2010 and January 2011. Baby-mother pairs were randomly selected from the Portuguese National Neonatal Screening Program. Eligible participants were mothers of three month old babies at the time of first interview, 16 years old or more, that were able to respond to the phone and understand and speak Portuguese. Exclusion criteria were: a gestational age lower than 36 weeks, birth weight lower than 2.500 kg, with severe malformations, needed intensive care and a formal contraindication to BF.

2.2 Survival analysis response variables and event of interest

The response variable *duration of any BF* is defined as the baby's age when was last breastfed. The event of interest is the *end of any BF*. Therefore, babies that were never BF are excluded from this analysis. The existence of competing events or multiple events was not considered.

The duration of any BF was calculated from the information gathered on the three time points. Code for variable construction is shown in Appendix C. At three months mothers were asked if they were presently BF, and for how long they had been BF. At six months only the information about current BF status was obtained. Therefore, at eighteen months it was necessary to contact all the mothers BF at three months - not only the ones BF at six months, but also the ones who stopped BF between the first and the second time points.

At three months, 494 mothers were successfully contacted. At six months there were 36 losses to follow up, also referred to as dropouts, so n=458. At eighteen months 262 contacts were attempted, corresponding to all the mothers BF at three months which had not been lost to follow up at six months, but there were 31 losses. Losses to follow up resulted in incomplete information about the response variable (censored observations) and missing values of covariates collected at the second time point. The decision to include any of these variables in model building would implicate the exclusion of these individuals, assuming no method of missing values imputation was used. Therefore, and also because most information was collected at three months, it was preferred to keep the maximum number of individuals in detriment of using any of the variables collected at six months. Among these, *post-partum depression* and *milk expression (3-6 months)*, are described in the bivariate analysis.

The duration obtained at eighteen months was cross-checked with the information of the two previous contacts and eleven inconsistencies were found (Table 2.1). Because most differences are small, they could be attributable to memory bias, in which case it would seem appropriate to consider the earliest information as correct. However, because this would result in censored observations close to

ID	BF at 3 months	BF at 6 months	BF duration at 18 months
	Duration at 18M lower than at 3M		
101	Yes	No	2
639	Yes	No	2.5
665	Yes	No	2
	Duration at 18M lower than at 6M		
349	Yes	Yes	4
	Duration at 18M higher than at 6M		
115	Yes	No	8
268	Yes	No	13
279	Yes	No	7.5
319	Yes	No	7
321	Yes	No	7
546	Yes	No	6.5
658	Yes	No	8

Table 2.1: Eleven cases were found in which BF duration reported at 18 months was inconsistent with previous answers.

event occurrence, and therefore result in informative censoring, the last information collected was considered for the analysis.

The response variable duration of exclusive BF is defined, according to the WHO definition, as the baby's age when other food or drinks besides breast milk, including water, was first introduced.

Exclusive BF was calculated from the information about food and water introduction collected at the three time points. In several cases different answers were obtained in different moments and the earliest information was preferred. In accordance to the first response variable, *duration of any BF*, only the mothers who breastfed exclusively for at least one week are considered to have breastfed exclusively.

2.3 Censoring

In this study censoring was caused by loss to follow-up (n=50) and if the mother was still BF at the time of last contact (n=42). All these cases are right censored because the event was not observed while the individual was under observation.

The assumption of independent censoring necessary for the validity of the inferences made using the methods described in chapter 3. Independent censoring means that the individuals censored at time t should be representative of all the individuals, with the same covariate values, that remain at risk at time t with respect to their survival experience (Kleinbaum & Klein, 2012; Rocha & Papoila, 2009).

Censoring at end of study, also called administrative censoring, is usually taken to be independent. However, losses to follow up could be related to factors associated with survival time.

In order to better describe censored individuals in this study, a linear logistic model was used to explain censored observations due to loss to follow up as a function of the explanatory variables included in the Cox regression model (Collett, 2003; Harrell, 2001).

2.4 Logistic regression response variable

The logistic regression attempts to model a binary outcome for BF, where success or failure are not determined by reaching a specific consensual duration, but instead are determined by the achievement of an individually set goal.

The response variable is binary with categories *achieved* and *failed*. The outcome is determined by the comparison of actual *BF duration* with *planned duration*, reported at the three months interview. A margin of half a month was applied in the comparison. As a result, mothers censored or who terminated BF within half a month or less of their intended duration were considered to have achieved their goal. R code for this variable is in Appendix C.

Besides a quantitative goal, intended duration was also collected as a qualitative variable, for example *until return to work* or *while the baby wants*.

Among the 473 mothers that initiated BF, 452 report a specific intended BF duration. However, due to censored observations, the total number of individuals for whom a specific outcome can be determined is reduced to 411. The 62 cases with an undetermined outcome include 16 cases with unspecified intended duration, 41 censored BF durations and 5 cases that are both.

Chapter 3

Cox Regression

Survival Analysis is an area of statistics that studies time until an event occurs, usually referred to as survival time. The event can be any event of interest that can happen to an individual such as death, relapse from a certain disease, or to an object, such as failure of a machine, etc. The survival time may be missing for some individuals due to various reasons like loss of contact or refusal (loss to follow-up), the end of the study before the event occurs, death by other causes, etc. In survival analysis these cases are very common and are described as censored observations. Censoring means there is some information about the individuals' survival time but its exact duration is unknown. The most common type is right-censoring and it occurs when the event of interest was not observed while the individual was under observation.

Let T be the random variable for survival time and t any specific value that T may take. In survival analysis the two major functions are the survival function and the hazard function. The survival function, $S(t)$, is defined as the probability that an individual will survive beyond a specific time point:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{3.1}$$

It can be seen from equation (3.1) that the survival function is the complement of the distribution function.

The hazard function, $h(t)$, can be interpreted as the instantaneous risk that the event will occur in an infinitesimal interval, between t and $t + dt$, conditional on the individual having survived until t:

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}. \tag{3.2}$$

Both the survival and hazard functions are related through formulas (3.3) and (3.4) and each one can be easily calculated from the other:

$$h(t) = -\frac{d \ln(S(t))}{dt}, \tag{3.3}$$

$$H(t) = \int_0^t h(u)\, du = -\ln S(t), \tag{3.4}$$

where $H(t)$ is the cumulative hazard function.

3.1 Exploratory techniques

The Kaplan-Meier (KM) estimator is used to estimate the survival function and its percentiles in the presence of censured observations. It divides the survival until time t into a sequence of independent events referring to each time interval until t, whose probabilities are conditional on the individuals at risk in each interval:

$$\widehat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}, \tag{3.5}$$

where n_i is the number of individuals at risk at time t_i and d_i is the number of events that occurred at the same time. Confidence intervals for the survival function estimate can be calculated using the Greenwood variance estimator:

$$\widehat{Var}(\widehat{S}(t)) = \widehat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \tag{3.6}$$

The median estimator for the survival function is defined as the smallest survival time for which the estimate of the survival function is lower or equal to 0.5:

$$\widehat{t}_{median} = min\{t_i : \widehat{S}(t_i) \leq 0.5\}. \tag{3.7}$$

Differences among levels of categorical variables or categorized continuous variables can be visually assessed through the plot of a stratified KM curve. Several tests can be applied to formally test the equivalence of the survival functions among levels (or groups), among which the log-rank and Peto-Peto tests (from here on referenced as Peto test). These tests use a chi-square test statistic that compares observed and expected cell counts and can be obtained from equation (3.8) applying different weights $w(t_i)$ to different survival times:

$$\frac{\left[\sum_i w(t_i)(d_{ij} - e_{ij})\right]^2}{Var\left[\sum_i w(t_i)(d_{ij} - e_{ij})\right]}, \tag{3.8}$$

where d_{ij} and e_{ij} are, respectively, the observed and expected number of events in group j at time i. The log-rank test applies the same weight to each failure time while the Peto test applies a weight that is a survival estimate similar to the KM survival estimate, $\tilde{\mathrm{S}}(t_i) = \prod\limits_{t_k \leq t_i} (1 - \frac{d_i}{n_i+1})$ calculated over all groups. Therefore, the Peto test places more weight on early failures (Kleinbaum & Klein, 2005; Rocha & Papoila, 2009).

3.2 Model formulation

Cox proportional hazards model is widely used for the analysis of medical data with censored observations (Cox, 1972). It has become a very popular technique because it is capable of providing good estimates of the regression coefficients without the need to specify a fully parametric functional form, given that its fundamental assumption is verified. The Cox regression models the hazard function as a product of a function of time, which is the baseline hazard, and a function of the covariates:

$$h(t|\boldsymbol{x}) = h_0(t)exp(\boldsymbol{\beta}'\boldsymbol{x}), \tag{3.9}$$

where $h_0(t)$ is the unspecified baseline hazard function, $\boldsymbol{x}$ is the vector of covariates and $\boldsymbol{\beta}$ is the vector of regression coefficients that represent the covariates effect on survival (Collett, 2003; Kleinbaum & Klein, 2005). The measure of effect is the hazard ratio (HR), which can be interpreted as the ratio between the hazard functions of individuals with different values of associated covariates:

$$\frac{h(t|\boldsymbol{x_1})}{h(t|\boldsymbol{x_2})} = \frac{h_0(t)exp(\boldsymbol{\beta' x_1})}{h_0(t)exp(\boldsymbol{\beta' x_2})} = exp(\boldsymbol{\beta'(x_1 - x_2)}). \tag{3.10}$$

Therefore, the exponential of each β can be interpreted as the hazard ratio between two individuals that differ in only that specific covariate. When $\beta < 0$, then $exp\,(\beta) < 1$, meaning that the variable lowers the hazard. As follows, when $\beta > 0$, then $exp\,(\beta) > 1$, the variable increases the hazard (Rocha & Papoila, 2009).

It is called a proportional hazards model because, according to equation (3.10), the hazard ratio between two different individuals is independent of time. The proportional hazards assumption implies that the effect of the covariates is constant over time.

Cox proposed the inference on the coefficients to be based on the following equation:

$$L\,(\beta) = \prod_{j=1}^{r} \frac{exp(\boldsymbol{\beta' x_{(j)}})}{\sum_{l \in R(t_{(j)})} exp(\boldsymbol{\beta' x_l})}, \tag{3.11}$$

where $R(t_j)$ is the number of individuals at risk just before t_j and $\boldsymbol{x}_{(j)}$ is the vector of covariates for the individual that dies at the jth ordered event time (Collett, 2003). Note that the numerator includes information about observed events while the denominator includes information about the individuals at risk, in a similar manner to the non parametric estimate of $S(t)$. This likelihood can be seen as a partial likelihood that allows inference on the coefficients without the specification of $h_0(t)$.

Due to the frequent presence of ties in real data, variants of the partial likelihood have to be used. In the R package *survival*, the Efron approximation is used by default (Therneau, 2013; Therneau & Grambsch, 2000).

Parameter estimates can be obtained by the maximization of the log-likelihood function. This is usually carried out using numerical methods, namely the Newton-Raphson procedure.

As sample size increases to infinity, the maximum likelihood estimators of β

are consistent, asymptotically normal and have covariance matrix $I\left(\beta\right)^{-1}$, where

$$I_{jk}\left(\beta\right) = -E\left(\frac{\partial^2 logL}{\partial\beta_j\partial\beta_k}\right). \tag{3.12}$$

$I\left(\beta\right)$ is the observed information matrix.

3.2.1 Covariate Selection

Based on the asymptotic normality of the maximum likelihood estimates, Wald's statistic can be used to test the null hypothesis $H_0 : \beta_j = 0$:

$$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim N(0,1). \tag{3.13}$$

The calculation of confidence intervals for the parameters is based on this statistic.

More generally, the importance of one or more parameters may be tested using the likelihood ratio test, based on the deviance statistic. The deviance statistic compares the likelihood functions of the adjusted model to the full model, which is a perfect fit to the data, and is calculated as follows:

$$-2 \times \ln(\hat{L}_{model}/\hat{L}_{full}) \sim \chi^2_{n-p-1}, \tag{3.14}$$

where $\hat{L}_{model}$ and $\hat{L}_{full}$ are, respectively, the maximized partial likelihoods of the model under consideration and the full model, n is the number of cases and p is the number of covariates. Intuitively, the lower the deviance, the better the model is adjusted.

When comparing two models, model 1 with p parameters nested in model 2 with $p+q$ parameters, the test statistic is as follows:

$$-2\ln(\hat{L_1}/\hat{L_2}) \sim \chi^2_q, \tag{3.15}$$

where $\hat{L}$ is the maximized likelihood function for each model. Under the null hypothesis that the extra q parameters in model 2 are all equal to zero, the test statistic has an asymptotic chi-square distribution with q degrees of freedom.

For large samples, the likelihood ratio test is asymptotically equivalent to Wald's statistic (Harrell, 2001). When this is not the case, the likelihood ratio test is preferred for being more robust (Carvalho *et al.*, 2011). However, model comparison using likelihood ratio requires that both models are estimated on the same data, meaning that the variables involved cannot have different missing values.

3.3 Model diagnostics

3.3.1 Residual analysis

Shoenfeld residuals

The proportional hazards assumption is verified through the graphical observation of scaled Shoenfeld residuals for each variable included and using the global and individual tests proposed by Therneau & Grambsch (2000).

For each individual, there is one residual for each covariate and these are only defined for observed event times. They can be interpreted as the covariate value for the individual that failed minus the expected value of the covariate for the individuals at risk, at the same time point. The residual for the jth covariate of the ith individual takes the form:

$$r_{ji} = \delta_i(x_{ji} - a_{ji}), \tag{3.16}$$

where δ_i is the indicator variable for censoring,

$$a_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)}{\sum_{l \in R(t_i)} exp(\hat{\boldsymbol{\beta}}' \boldsymbol{x}_l)}, \tag{3.17}$$

$R(t_i)$ is the set of individuals at risk at time t_i (Collett, 2003). For large samples their expected value is zero. The r_{ji} are uncorrelated and, therefore, when plotted against event time, they should exhibit a random pattern around zero.

The scaled Schoenfeld residuals are a transformation of the original residuals that makes them more useful in the detection of departure from the proportional hazards assumption. Considering $\mathbf{r}_i$ as the vector of Schoenfeld residuals for the ith individual, the scaled residuals will be the components of the vector:

$$\mathbf{r}_i^* = r \times var(\hat{\boldsymbol{\beta}})\mathbf{r}_i \tag{3.18}$$

where r is the total number of events and $var(\hat{\boldsymbol{\beta}})$ is variance-covariance matrix of the parameters estimates for the model. It has been shown that the expected value of the residual for the ith individual and the jth covariate is given by

$$E(r_{ji}^*) \approx \beta_j(t_i) - \hat{\beta}_j \tag{3.19}$$

where $\beta_j(t)$ can be seen as a time-varying coefficient of the X_j covariate and $\hat{\beta}_j$ is the estimated coefficient. Therefore, the plot of $r_{ji}^* + \hat{\beta}_j$ against time will reveal the form of the time-varying coefficient. If the proportional hazards assumption is verified, the adjusted smoothed curve will correspond to an horizontal line.

Martingale residuals

Martingale residuals are used to assess the functional form of continuous covariates and the existence of individuals poorly adjusted by the model. The martingale residual associated with the ith individual is given by:

$$M_i = \delta_i - exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_i)\hat{H}_0(t_i) \tag{3.20}$$

where $\hat{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function. These residuals take values between $-\infty$ and unity and are unevenly distributed around zero. In large samples, they are uncorrelated and have an expected value of zero (Collett, 2003). The plot of martingale residuals of the null model against the observed values of a continuous covariate should reveal its functional form. The adequacy of the transformation used would be confirmed by a straight line for the smoothed curve. The plot of martingale residuals against the individuals' index, ranked by survival time, will reveal the existence of outliers, individuals for whom the actual survival time is much smaller or larger than predicted by the model.

Deviance residuals

Another type of residuals used for the identification of outliers are the deviance residuals. These are a transformation of martingale residuals that are more symmetrically distributed around zero:

$$D_i = sgn(\hat{M}_i)\sqrt{-2 \times (\hat{M}_i + \delta_i \ln(\delta_i - \hat{M}_i))} \tag{3.21}$$

where the sign function ensures that the deviance residual for the ith individual takes the same sign as the correspondent martingale residual. These residuals can also be interpreted as components of the deviance statistic (Eq. 3.14), in a way that a large residual corresponds to an observation poorly adjusted by the model (Collett, 2003).

Score residuals

Score residuals are useful in the detection of influential observations. The influence of an observation i on a parameter estimate $\hat{\beta}_j$ can be measured by the change in the estimate of the coefficient caused by the omission of the observation:

$$\hat{\beta}_j - \hat{\beta}_{j(-i)}, \tag{3.22}$$

where $\hat{\beta}_{j(-i)}$ is the coefficient estimated without the ith observation. The actual difference can only be calculated by fitting the same model to the n datasets, omitting one observation at a time, which can be computationally intensive. However, an approximation to the change in the estimate is based on the score residuals for the fitted model. This quantity is called *delta-beta*, denoted by $\Delta_i\hat{\beta}_j$, and is given by the jth component of the vector

$$\boldsymbol{r}'_{Si} var(\hat{\boldsymbol{\beta}}), \tag{3.23}$$

where $\boldsymbol{r}_{Si}$ is the vector of score residuals for the ith individual and $var(\hat{\boldsymbol{\beta}})$ is the variance-covariance matrix of the parameter estimates. The delta-beta divided by the standard error of $\hat{\beta}_j$ will give a standardized delta-beta that can be used to assess the significance of the change (Collett, 2003).

The score residual for the ith individual and jth coefficient may be written in the form:

$$r_{Sji} = r_{ji} + exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_i) \sum_{t_r \leq t_i} \frac{(\hat{a}_{jr} - x_{ji})\delta_r}{\sum_{l \in R(t_r)} exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x}_l)}, \tag{3.24}$$

where r_{ji} is the Shoenfeld residual and x_{ji} is the jth covariate value for the ith individual and $R(t_r)$ is the risk set at time t_r.

3.3.2 Model adjustment

In survival analysis, in a similar manner to a generalized linear model, quality of fit can be given by the proportion of total variability in the response that is explained by the covariates. It is based on an R^2 measure, among which is the Cox and Snell's R^2:

$$R^2 = 1 - (\hat{L}_{null}/\hat{L}_{model})^{2/n}, \tag{3.25}$$

where $\hat{L}_{null}$ is the maximized partial likelihood of the null model. The proportion of total variability explained can be calculated as the ratio of the R^2 of the model and the R^2 of the full model (Carvalho *et al.*, 2011).

Another measure of fit used is the concordance probability, which measures the discriminatory power of the model. It is analogous to the area under the ROC curve in logistic regression (Carvalho *et al.*, 2011). Both measures are easily obtained in the model's output.

Chapter 4

Application - Cox Regression

The association of independent variables with the *duration of any BF* is explored through KM curves and log-rank and Peto tests. Variables with a p-value below 0.20 for at least one of the tests are considered for inclusion in the model.

These variables are grouped as demographic, biological, social and psychological determinants. Model building took into account the conceptualization of the phenomenon under study by establishing a hierarchy in the groups of determinants (Thulier & Mercer, 2009). This hierarchy is based on the causal and temporal relations among the different levels and, in this manner, an attempt is made to adjust for potential confounders (Fuchs *et al.*, 1996; Victora *et al.*, 1997).

Inside each block or level of modelling, variables are selected according to a stepwise method described in Turkman & Silva (2000). Variables are included if the p-value is lower than $p_E = 0.20$ and excluded if the p-value is greater than $p_R = 0.25$. The likelihood ratio test is used for model comparison. The variables *maternal age* and *education* are kept in the model independently of statistical significance. These demographic variables are consistently identified in the literature as BF determinants and are potential confounders for other determinants.

4.1 Exploratory Analysis

The characteristics of the study participants are described in table 4.1. Mothers were between 17 and 45 years old, with a mean age of $30.8(\pm 5, 2)$ years.

	N	%
Maternal Age		
≤ 20	15	3.0
21-25	72	14.6
26-30	135	27.4
31-35	173	35.1
>35	98	19.9
Years of education		
≤ 9	118	24.0
10-12	144	29.3
≥ 12	229	46.7
Mother's country of origin		
Portugal	415	84.2
PALOP	27	5.5
Brazil	22	4.5
Other	29	5.9
Region of residence		
Lisbon	164	33.2
Center	92	18.6
Algarve and Alentejo	45	9.1
North	163	33.0
Madeira and Azores	30	6.1
Parity		
Primiparous	267	54.0
Multiparous	227	46.0
Method of delivery		
Normal	306	61.9
Cesarean	174	35.2
Other	14	2.8

Table 4.1: Main characteristics of study sample participants.

Nearly half of them had a higher or post-secondary educational level. Immigrants accounted for almost 16% of the mothers, the majority of which were from Portuguese-speaking African countries (PALOP). The proportion of first time mothers was 54% and about 70% had a normal delivery.

Only 21 mothers were not BF at the end of the first week after birth, so BF initiation rate was 95.7%. The remaining 473 mothers that initiated BF are considered for the analysis of BF duration. Among these there are 370 observed survival times and 92 censored survival times. Their distribution is shown in Figure 4.1. As explained in Section 2.2, 50 observations (10.5%) are censored because of loss to follow up and 42 mothers (8.8%) were still breastfeeding at the time they were last contacted.

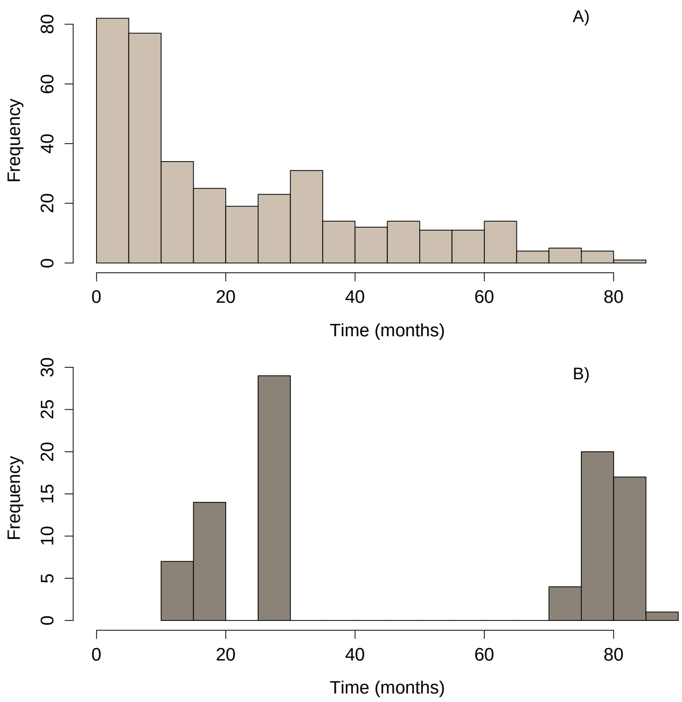


Figure 4.1: Histogram of observed (A) and censored (B) survival times for *BF duration*.

The Kaplan-Meier method was used to estimate the survival function and the median survival time for BF duration. The estimated survival function for both *any BF* and *exclusive BF* are shown in Figure 4.2. The median survival time for *any BF* was 5 months (n=473). The median survival time for *exclusive BF* was 9 weeks, or 2 months (n=455).

Maternal age, *BSES* and *motivation scale* are categorized for the purpose of exploratory analysis only and are used as continuous in the model. The distribution of *maternal age* and *BSES* are shown in Figure 4.3.

The exploratory analysis revealed several variables that significantly affect the duration of BF. These are described as follows.

The following categories of demographic variables were found to be positively associated with BF duration: higher maternal age, being married or partnered, being an immigrant, belonging to an ethnic group other than white, having a secondary or higher education and being part of a larger family($\geq$ 4 persons).

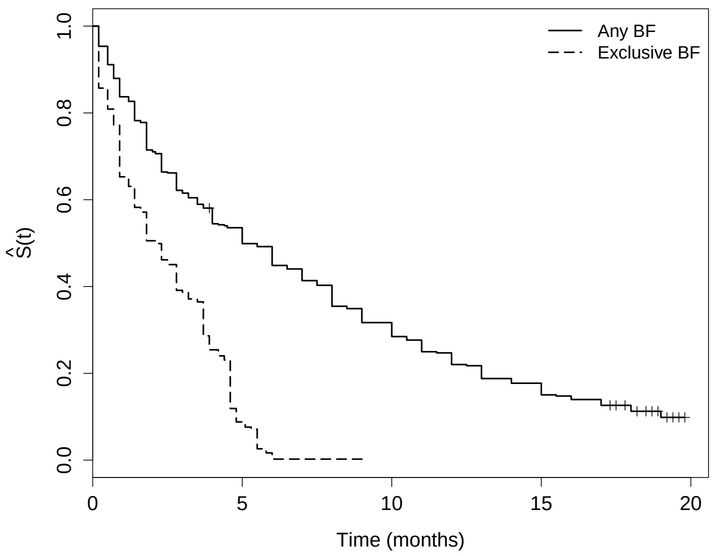


Figure 4.2: Kaplan-Meier estimate of the survival function for A-*any BF* and B-*exclusive BF*.

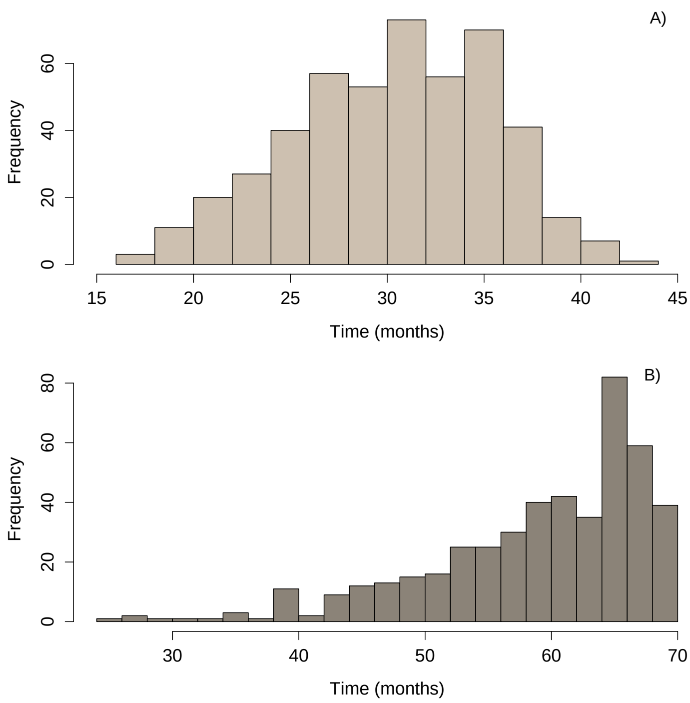


Figure 4.3: Histogram of *maternal age* (A) and *BSES* (B) (n=473).

Variable	N	Median	Log Rank	Peto
Maternal Age				
≤ 25	84	3.2		
26-34	256	5.5		
≥ 35	133	8.0	0.0262	0.0050
Marital status				
Has a partner	406	6.0		
No partner	65	3.7	0.1750	0.0870
Country of origin				
Portugal	395	5.0		
Other	77	10.5	0.0003	$<$0.0001
Education				
Basic	113	3.5		
Secondary or higher	359	6.0	0.0280	0.0009
Family size				
≤ 3	176	5.0		
≥ 4	204	6.0	0.1620	0.2610

Table 4.2: Median survival time and p-value for log-rank and Peto tests for demographic variables.

No differences in the probability of survival were found between secondary and higher educational levels and so these categories were merged. These variables are summarized in Table 4.2 and their KM curves are plotted in Figure 4.4.

Among the biological factors, having had or having breastfed other children have a positive effect on duration while smoking during pregnancy, having experienced complications (until three months) or having been diagnosed with postpartum depression (until six months) had a negative effect. Differences were also found among methods of delivery, with cesarean and normal birth with epidural being associated with lower durations of BF than normal birth without epidural. These variables are shown on Table 4.3 and Figure 4.5.

Social variables include family support to breastfeeding as well as the support from health care system. The following maternity practices had a positive effect on the duration of BF: first BF within 30 minutes of birth, BF on demand, being taught how to breastfeed and not introducing formula. Attending a BF support group at the health center in the postpartum period, having been advised to breastfeed and the father being in favour of BF were also associated with a longer duration. Social variables are shown on Table 4.4 and Figure 4.6.

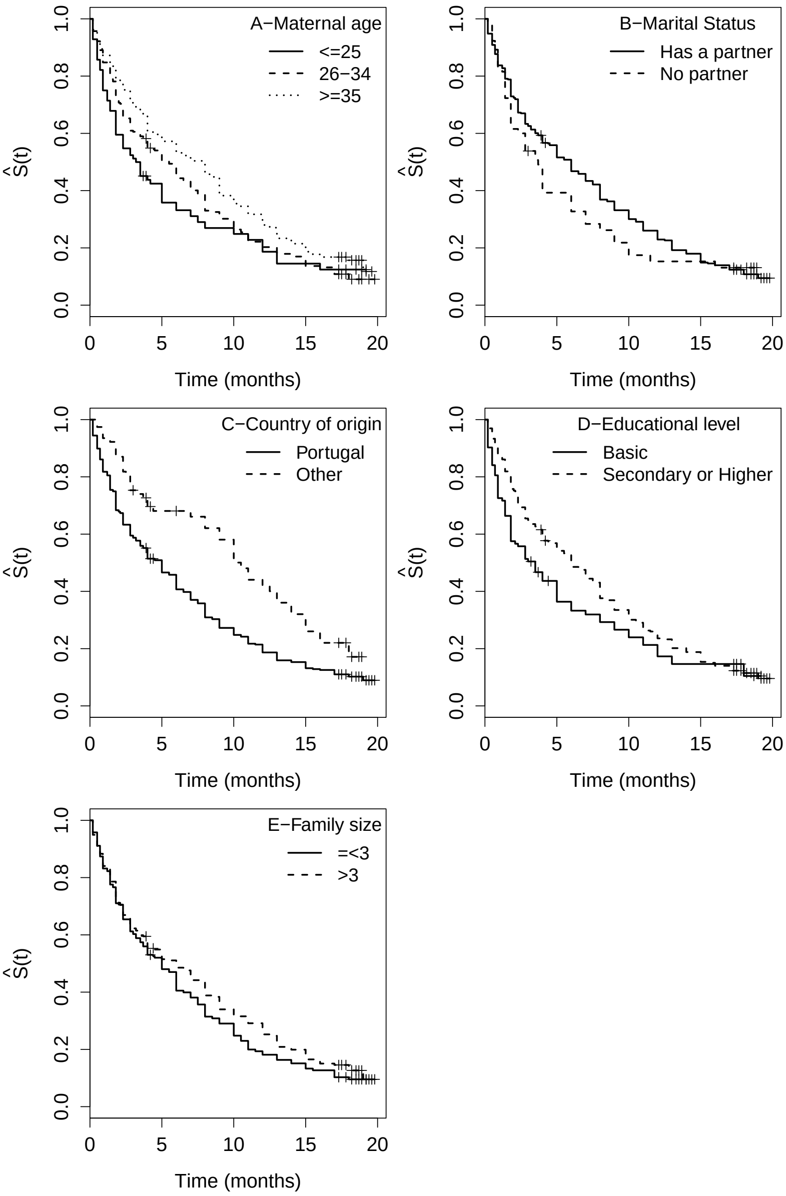


Figure 4.4: Kaplan-Meier estimate of the survival function for BF duration, stratified by demographic variables: A-*Maternal age*, B-*Marital Status*, C-*Country of origin*, D-*Education*, E-*Family size*.

Variable	N	Median	Log Rank	Peto
Previous BF experience				
No	261	5.0		
Yes	212	6.0	0.0193	0.0109
Smoked during pregnancy				
Yes	57	2.8		
No	416	6.0	0.0039	0.0058
Method of delivery				
Cesarean	165	6.0		
Normal, epidural	226	4.0		
Normal, no epidural	82	8.0	0.0039	0.0050
Any BF Complications (first 3M)				
No	271	6.0		
Yes	200	5.0	0.2740	0.1290
Postpartum depression (first 6M)				
No	413	6.0		
Yes	25	2.0	0.0042	0.0004
Blocked ducts				
No	426	6.0		
Yes	46	3.6	0.1580	0.0889
Mastitis				
No	446	4.5		
Yes	26	2.0	0.0674	0.119
Cracked Nipples				
No	375	6.0		
Yes	97	4.5	0.1180	0.382

Table 4.3: Median survival time and p-value for log-rank and Peto tests for biological variables.

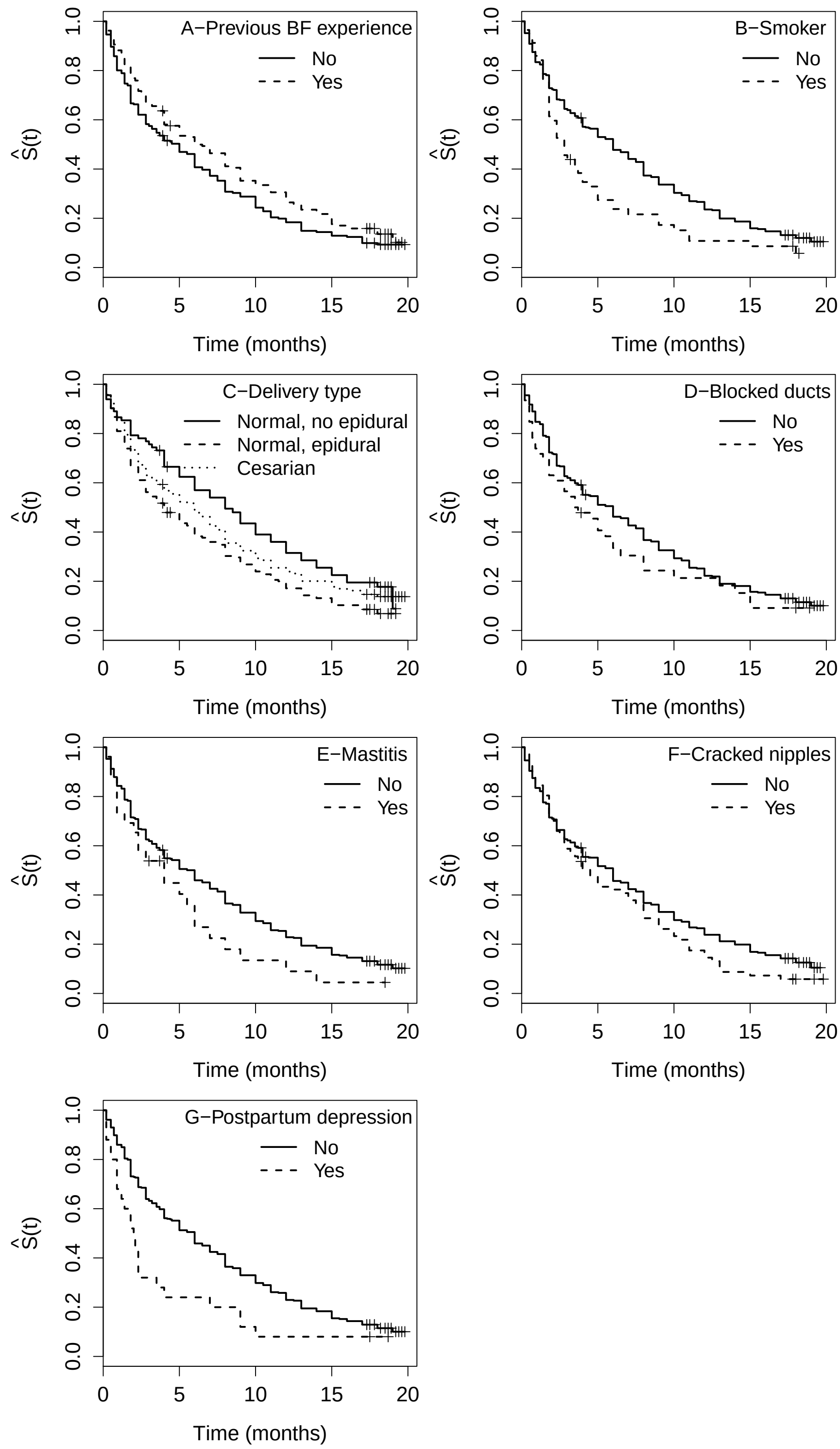


Figure 4.5: Kaplan-Meier estimate of the survival function for BF duration, stratified by biological variables: A-*Previous BF experience*, B-*Smoked during pregnancy*, C-*Method of delivery*, D-*Blocked ducts*, E-*Mastitis*, F-*Cracked nipples*, G-*Postpartum depression*

Variable	N	Median	Log Rank	Peto
Mother was BF				
No/Does not know	93	3.7		
Yes	378	6.0	0.0029	0.0057
Was advised to BF				
No	154	4.0		
Yes	318	6.0	0.3580	0.1570
Father's opinion of BF				
In favour	447	6.0		
Other	26	2.1	0.0283	0.0054
BF within 30 minutes (maternity)				
No	251	5.0		
Yes	196	7.0	0.1400	0.0315
BF on demand (maternity)				
No	41	2.8		
Yes	428	6.0	0.0635	0.0246
Formula introduction (maternity)				
No	294	6.0		
Yes	175	3.0	0.0014	0.0001
Was taught how to BF (maternity)				
No	117	6.0		
Yes	354	5.0	0.2200	0.1910
Attended BF support group (health center)				
No	420	5.0		
Yes	51	7.0	0.0159	0.0188

Table 4.4: Median survival time and p-value for log-rank and Peto tests for social variables.

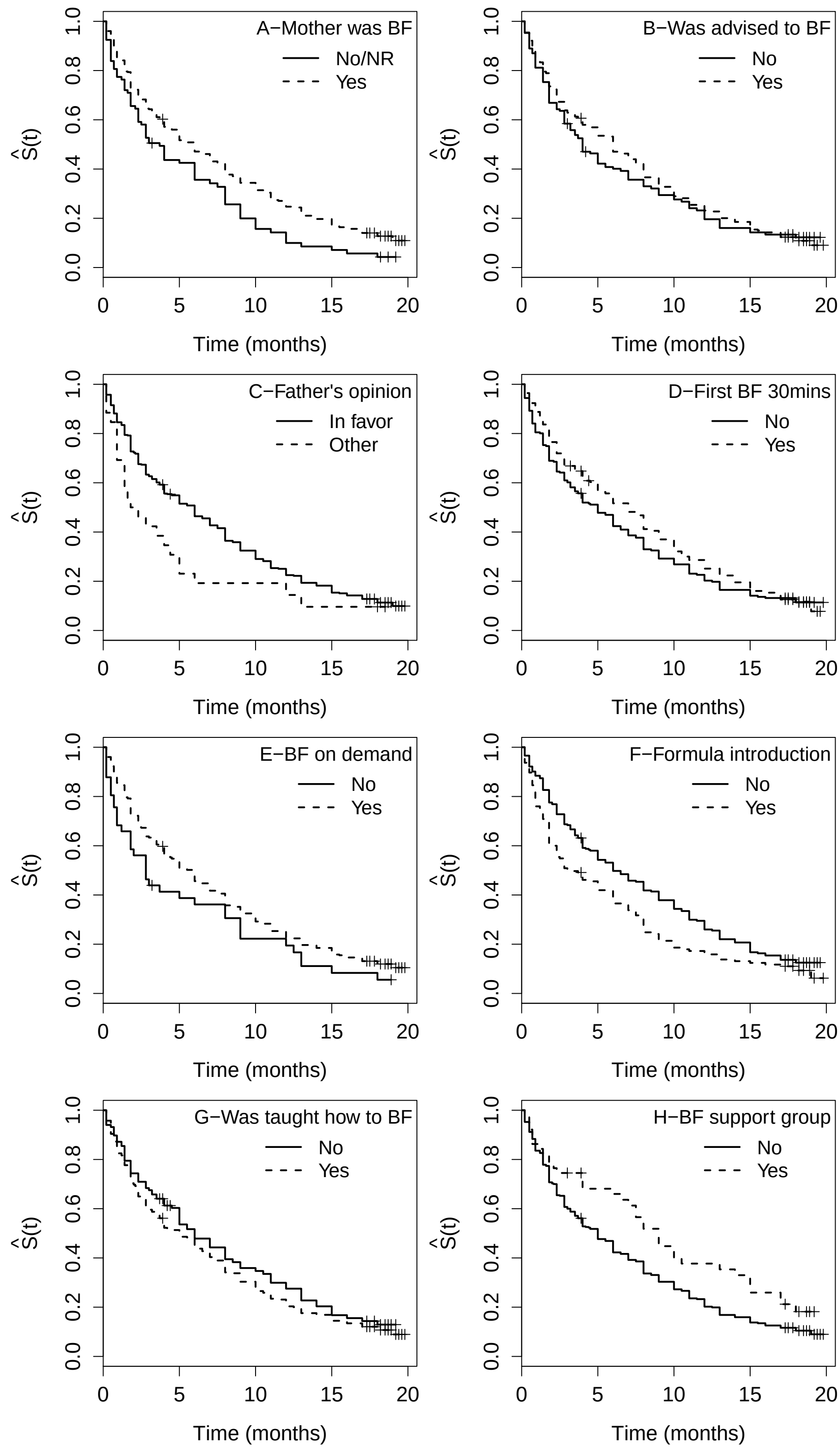


Figure 4.6: Kaplan-Meier estimate of the survival function for BF duration, stratified by social variables: A-*Mother was BF*, B-*Was advised to BF*, C-*Father's opinion of BF*, D-*BF within 30 minutes*, E-*BF on demand*, F-*Formula introduction*, G-*Was taught how to BF*, H-*Attended BF support group.*

Psychological variables are related to the mothers' motivation, confidence and knowledge about BF. Several variables related to the mothers' motivation were associated with longer durations such as the mother having been breastfed herself, having made the decision to BF before birth, intending to breastfed for longer than six months, being highly motivated before birth and in the first week after birth, recognizing personal benefits among the reasons to breastfeed, and recognizing that BF confers nutritional benefits, strengthens the mother-baby relationship and is practical. The mothers' confidence is measured in the *breastfeeding self efficacy scale* (*BSES*) score as well as in the variable *ask for help*. The use of BF aids and BF facilitating techniques can be seen as a reflection of the mothers' knowledge about BF. Using more facilitating techniques was associated with a longer duration of BF. Pacifier introduction before one month of age and milk expression during the first three months had a negative effect on the duration of BF. On the other hand, the practice of milk expression between three and six months of age appears to have a positive effect. Psychological variables are shown in Figures 4.7 and 4.8.

Demographic and biological variables identified are much in accordance with other studies. However, other variables commonly identified in the literature as influencing BF duration or found to be associated with BF duration in similar studies were not found to be significant at the level of significance used. These include the working status of the mother before delivery (log-rank $p = 0.838$), attendance of prepartum classes ($p = 0.647$) and grandmother's opinion about BF ($p = 0.448$).

Other criteria for variable selection

Some variables in the dataset corresponded to different categorizations of the same determinant or contained essentially the same information. Even though they are associated with BF duration, it is not necessary to consider all for modelling. The choice among them was based, first of all, on the number of missing values and, secondly, on the likelihood ratio test for the simple model with each covariate.

Information about maternal education was collected both as continuous (years of education) and as categorical. In this case the categorical variable is preferred

Variable	N	Median	Log Rank	Peto
Motivation				
Decided to BF				
Before birth	452	5.5		
After birth/Does not know	21	2.8	0.0880	0.0580
Intended BF duration				
$\leq$ 6 months/undecided	284	6.0		
>6 months	188	4.0	0.0526	0.0624
Motivation scale (1st week after delivery)				
1-4	149	3.7		
5	324	7.0	0.0026	0.0042
Personal benefits (motivation to BF)				
No	319	5.0		
Yes	154	6.0	0.2250	0.1810
Confidence				
BSES score at 3 months				
[14,57)	138	1.8		
[57,65)	147	5.0		
[65,70]	180	8.0	<0.0001	<0.0001
Ask for help				
No	349	6.0		
Yes	123	2.8	0.0002	<0.0001
Knowledge				
Nutritional benefits				
No	199	5.0		
Yes	274	6.0	0.3300	0.1280
Strengthens the mother-baby relation				
No	261	4.0		
Yes	212	6.0	0.0120	0.0088
Is practical				
No	385	5.0		
Yes	88	7.5	0.1240	0.0214
No. BF techniques used				
0-1	212	4.0		
2-3	261	6.5	0.0056	0.0012
Pacifier at age 1 month				
No	125	10.5		
Yes	348	3.7	<0.0001	<0.0001
Milk expression (0-3 months)				
No	179	8.0		
Yes	294	4.0	0.0025	0.0016
Milk expression (3-6 months)				
No	163	2.8		
Yes	114	11.0	0.0032	<0.0001

Table 4.5: Median survival time and p-value for log-rank and Peto tests for psychological variables.

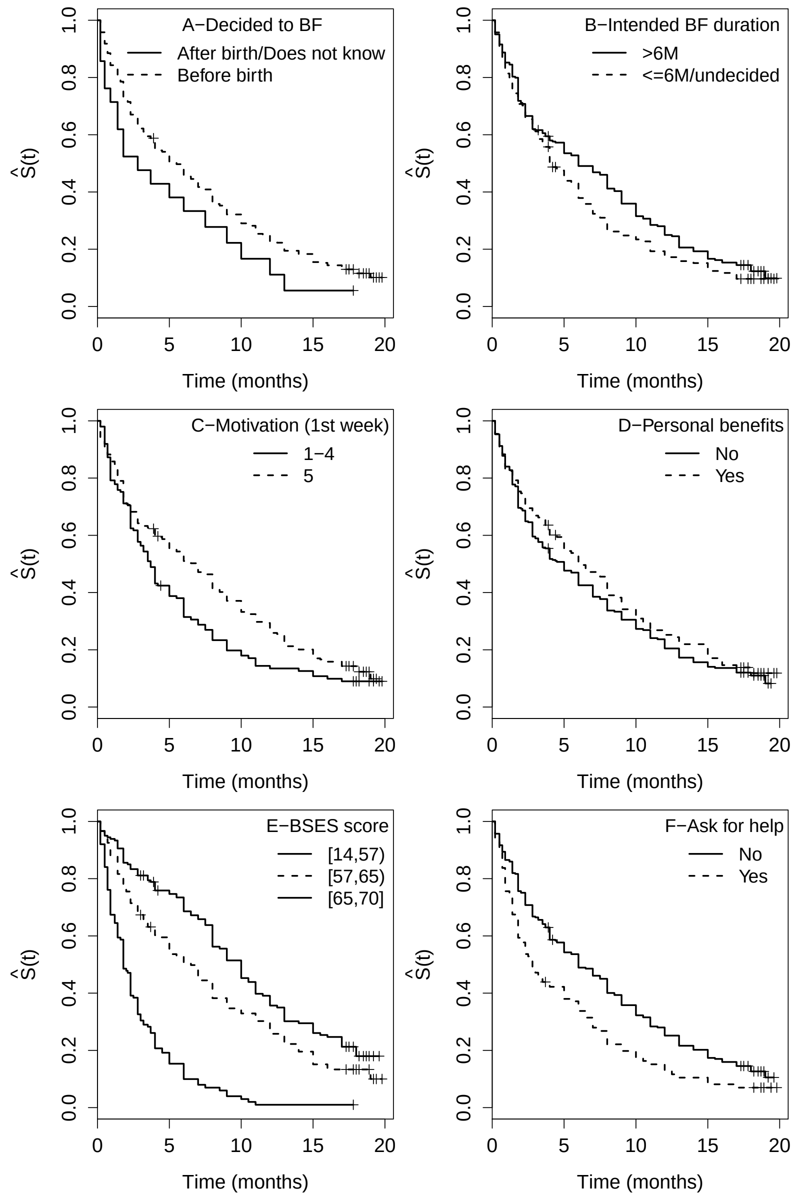


Figure 4.7: Kaplan-Meier estimate of the survival function for BF duration, stratified by psychological variables: A-*Decided to BF*, B-*Intended BF duration*, C-*Motivation scale* (1st week after delivery), D-*Personal benefits* in decision to BF, E-*BSES* score, F-Frequently *asks for help* about BF.

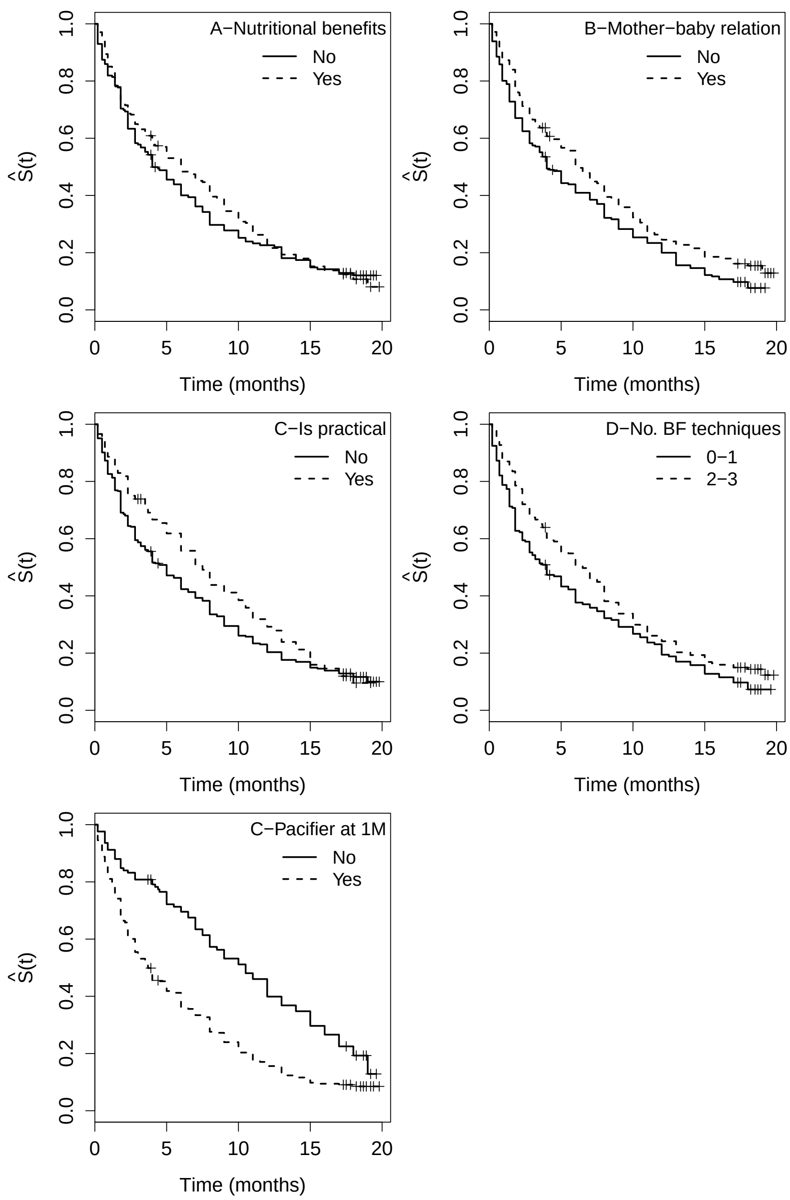


Figure 4.8: Kaplan-Meier estimate of the survival function for BF duration, stratified by psychological variables: A-*Nutritional benefits*, B-*Strengthens the mother-baby relationship*, C-*Is practical*, D-Uses *pacifier at age 1 month.*

because it has a lower number of missing values. The variable *country of origin* is also preferred over *ethnic group* because of a lower number of missing values. The variables *previous BF experience* and *parity* are preferred over *previous positive BF experience* because of the number of missing values, and among them *previous BF experience* was selected based on the likelihood ratio test (p=0.0127 versus p=0.0399).The motivation scale for the first week after delivery is preferred over the pre-partum based on the likelihood ratio test (p=0.0037 versus p=0.1480). For simplicity, only the chosen variables are shown in figures and tables.

The variables *postpartum depression* and *BF within 30 minutes* were excluded from model building because of a high number of missing values. The variable *ask for help* was also excluded from modelling because it is closer to being a consequence of the response variable than a predictor.

BF aids

The variables related to BF aids such as pacifier and milk bottle, as well as the practice of milk expression are time-dependent in nature but were collected in different forms. For pacifier and bottle there is information about the time they were first introduced. This information is complete at three months but there are several missing values at six months because of dropouts and because the question was not applied.

The question about milk bottle introduction referred specifically to its use with breast milk and not formula. However, it was misinterpreted and among the 366 cases of bottle introduction until six months, 42% refer introducing formula on the same week. Because of reverse-causality bias this variable will not be considered.

Pacifier introduction was categorized as having been introduced before one month of age or otherwise because its introduction is not recommended until BF is established, at 3 or 4 weeks of age (Village, 2012). This last category includes all the individuals for which we know the time of introduction and also the ones for whom that information is missing after three months.

The practice of milk expression was questioned at three months for all the women that ever BF and at six months for the women BF at the previous interview, independently of the previous answer. It is a time-dependent covariate

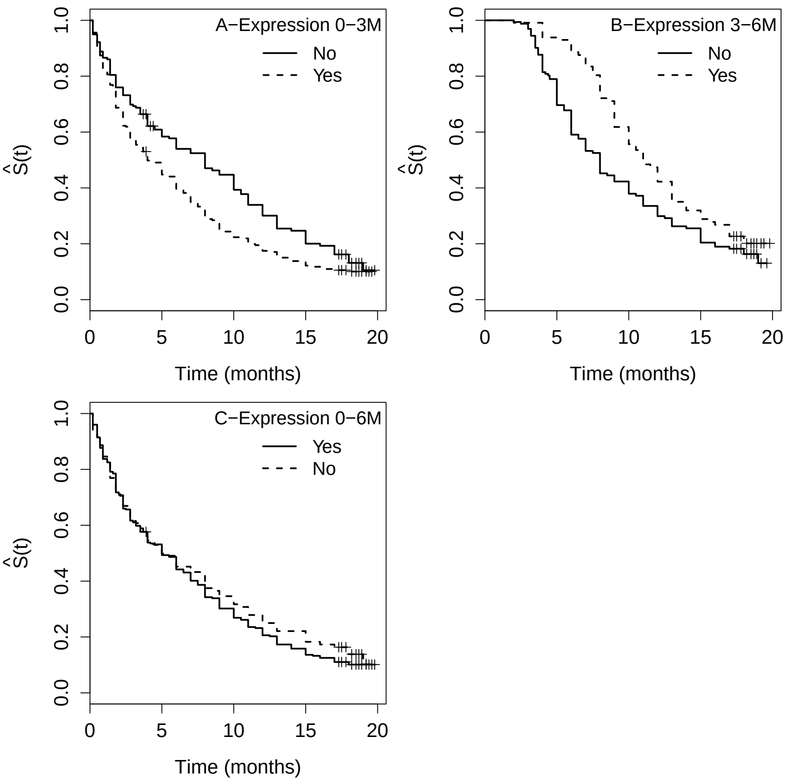


Figure 4.9: Kaplan-Meier estimate of the survival function for BF duration, stratified by the practice of milk expression: A- between 0 and 3 months, B-between 3 and 6 months (for the ones BF at 3 months, n=277), C-between 0 and 6 months.

that, as mentioned, appears to have a time-varying effect on BF survival and its overall effect is null (Figure 4.9). Although the use of milk expression in the first three months could be considered for modelling, it would not express the complete information about how the variable affects BF duration. It was, therefore, not considered for modelling at all.

4.2 Model Building

Tables 4.6, 4.7, 4.8 and 4.9 show the p-values associated with variable selection for each step of the four levels of the hierarchic model. The initial model considered at the beginning of level 1 included *maternal age* and *education*. In decision step

1 of level 1 (first line in Table 4.6), *country* is included with $p < 0.001$, which is lower than $p_E = 0.20$. In step 2, no variable is included because the lowest p-value, 0.295 for *marital status*, is above the inclusion criteria. At each step, the p-values for exclusion are below the horizontal line and the p-values for inclusion are above. The highest exclusion p-values at each level is marked with an asterisk. After the end of each level, the variables included are no longer considered for exclusion. Variable selection included a total of 17 steps and variables included were never excluded in a subsequent step. The model chosen at the end of level 4 is described in Table 4.10.

decision step	Country	Marital Status	Family size
1	<0.001	0.417	0.435
2	<0.001	0.295	0.680

Table 4.6: Variable selection in level 1.

decision step	Smoker	Delivery	Cracked nipples	Mastitis	Blocked ducts	Previous BF experience
1	0.012	0.019	0.089	0.034	0.181	0.229
2	0.012	0.022	0.116	0.048	0.292	0.263
3	0.015	0.022*	0.059	0.062	0.353	0.333
4	0.019	0.013	0.059*	0.062	0.360	0.447
5	0.027	0.016	0.059	0.062*	0.372	0.415

Table 4.7: Variable selection in level 2.

decision step	Mother was BF	Support group	Formula Introd.	BF on demand	Father's opinion	BF taught	BF advice
1	<0.001	0.011	0.005	0.096	0.172	0.318	0.665
2	<0.001	0.004	0.010	0.045	0.145	0.295	0.573
3	<0.001	0.004*	0.010	0.083	0.223	0.285	0.797
4	<0.001	0.004	0.010*	0.044	0.455	0.436	0.768
5	<0.001	0.007	0.005	0.044*	0.412	0.531	0.642

Table 4.8: Variable selection in level 3.

decision step	BSES	Pacifier 1M	No. BF practices	Mother-baby relation	Nutritional	Practical	Motivation 1stweek	Decision timing	Planned duration	Personal benefits
1	<0.001	<0.001	0.021	0.049	0.224	0.243	0.009	0.409	0.005	0.486
2	<0.001	0.014	0.021	0.045	0.281	0.346	0.159	0.324	0.746	0.760
3	<0.001	0.014*	0.020	0.036	0.325	0.385	0.199	0.387	0.656	0.640
4	<0.001	0.014	0.020*	0.061	0.287	0.440	0.351	0.368	0.472	0.793
5	<0.001	0.012	0.034	0.061*	0.268	0.326	0.372	0.453	0.457	0.748

Table 4.9: Variable selection in level 4.

4.3 Model Diagnostics

4.3.1 Residual analysis

The smoothed curve of the Schoenfeld residual plots (Figures 4.10 and 4.11) did not reveal a systematic departure from an horizontal line for any of the variables in the model. However, the linear correlation test reveals a possible violation of the proportional hazards assumption for the variable *education*. The correspondent residual plot shows that for a short period at the beginning of the time scale the estimated model coefficient is not contained inside the confidence interval of the curve, although there is not an overall growth trend for $\beta(t)$.

The plot of martingale residuals of the null model against the two continuous variables in Figure 4.12 suggests that *maternal age* might have a non linear effect. Given the shape of the curve, this apparent non-linearity was addressed using a P-*spline* function. However, the output for this model showed that the linear term for *maternal age* was significant (p=0.067) and that the non-linear term was not significant (p=0.320). The likelihood ratio test confirmed the adequacy of the simpler model (p=0.9341).

Martingale and deviance residuals (Figure 4.13) reveal some observations are poorly adjusted by the current model. There are 35 observations (7.8%) with values of deviance residuals outside the range between -2 and 2. The ten individuals with values outside the range (-2.5,2.5) are identified in Figure 4.13 B.

The analysis of standardized delta-beta values, shown in Figures 4.14 and 4.15 reveals the presence of some influencial observations. These are identified in the figure for the variables *maternal age* and *education*.

Individuals 2 and 257 are simultaneously outliers and influential for the parameter estimates of *maternal age*, *education* and *mother was BF*. It can be seen

	Hazard ratio	95% CI	p-value $H_0 : \beta_j$=0
	Demographic		
Maternal age	0.980	0.959-1.002	0.0743
Level of education			
Basic	1		
Secondary or Higher	0.793	0.609-1.031	0.0835
Country of origin			
Portugal	1		
Other	0.680	0.493-0.938	0.0189
	Biological		
Smoked during pregnancy			
No	1		
Yes	1.418	1.029-1.954	0.0329
Method of delivery			
Normal, no epidural	1		
Normal, with epidural	1.238	0.910-1.684	0.1737
Cesarean	1.119	0.813-1.541	0.4915
Cracked nipples			
No	1		
Yes	1.086	0.834-1.413	0.5400
Mastitis			
No	1		
Yes	1.696	1.093-2.633	0.0185
	Social		
Mother was breastfed			
No	1		
Yes	0.727	0.559-0.947	0.0180
Attend BF support group			
No	1		
Yes	0.756	0.534-1.070	0.1147
Formula introduction (maternity)			
No	1		
Yes	1.060	0.845-1.330	0.6141
BF on demand (maternity)			
No	1		
Yes	0.678	0.471-0.975	0.0359
	Psychological		
BSES score	0.930	0.918-0.943	$<$0.0001
Pacifier at 1 month			
No	1		
Yes	1.391	1.071-1.807	0.0135
No. of BF techniques used			
0-1	1		
2-3	0.795	0.643-0.983	0.0339
Mother-baby relationship benefits			
No	1		
Yes	0.808	0.646-1.011	0.0623

Table 4.10: Cox regression model output for the duration of any BF. Complete case analysis for n=449 individuals, including 84 censored observations.

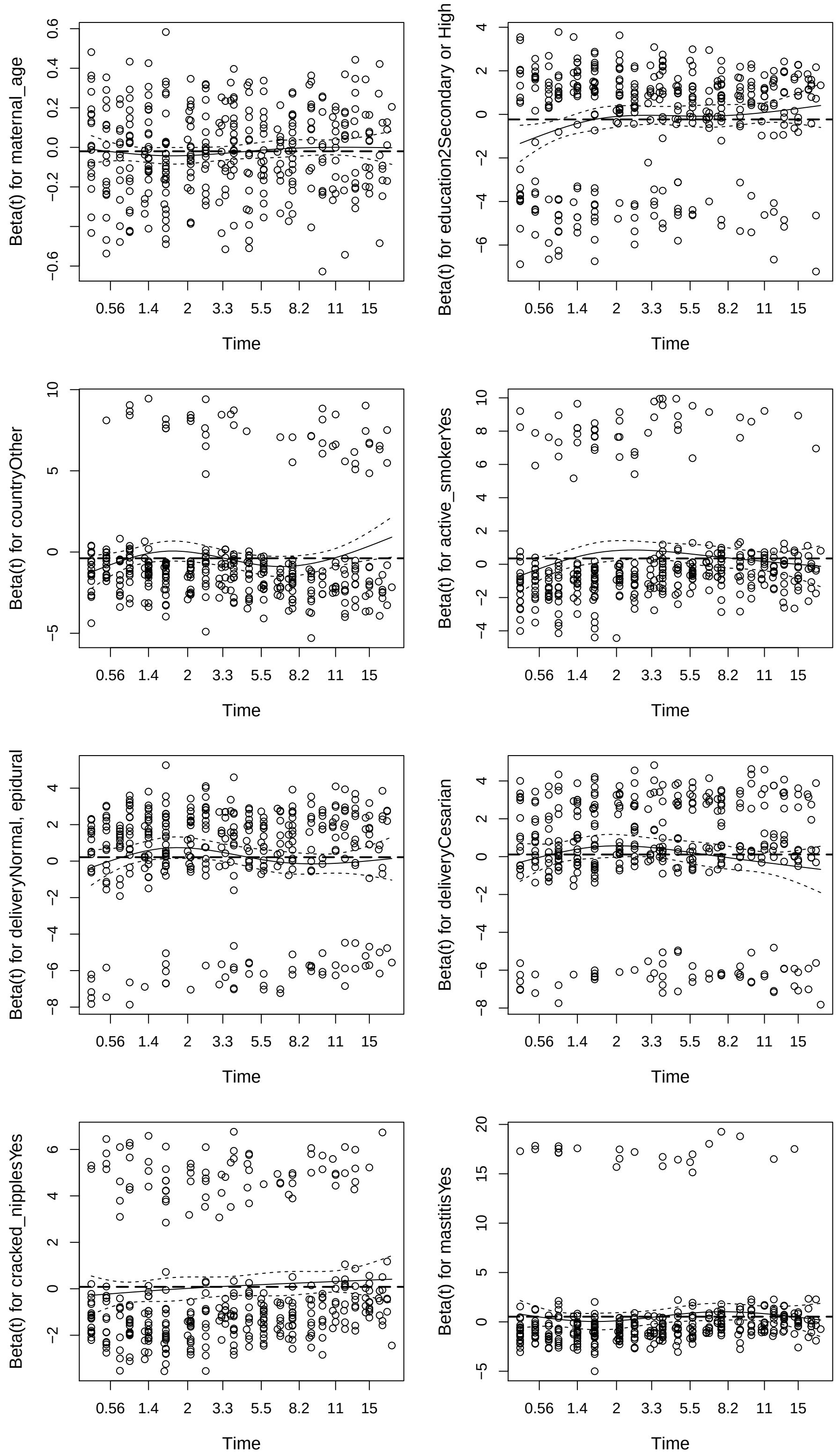


Figure 4.10: Scaled Schoenfeld residuals for the current model.

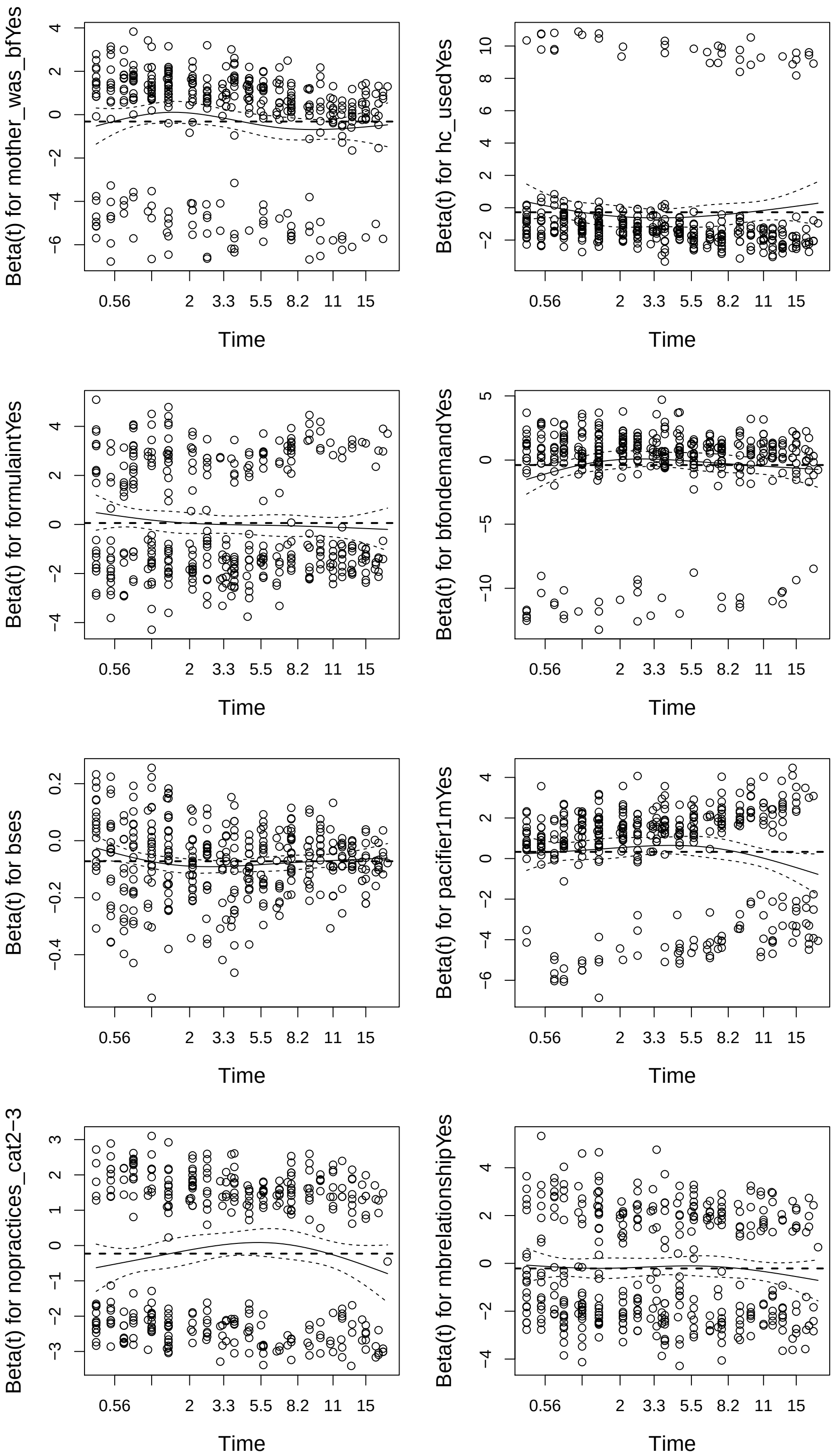


Figure 4.11: Scaled Schoenfeld residuals for the current model.

	rho	chisq	p-value
Maternal age	0.0503	0.9676	0.3253
Education Secondary/Higher	0.1417	7.3519	0.0067
Country Other	0.0452	0.7740	0.3790
Smoked during pregnancy Yes	0.0175	0.1166	0.7327
Method of delivery Epidural	-0.0397	0.5836	0.4449
Method of delivery Cesarian	-0.0600	1.3509	0.2451
Cracked nipples Yes	0.0728	2.1820	0.1396
Mastitis Yes	0.0343	0.4648	0.4954
Mother was BF Yes	-0.0690	1.8616	0.1724
Used support group Yes	-0.0098	0.0364	0.8487
Formula introduction Yes	-0.0730	2.1235	0.1451
BF on demand Yes	0.0334	0.4456	0.5044
BSES score	-0.0435	0.6965	0.4040
Pacifier at age 1 month Yes	-0.0655	1.6069	0.2049
No. BF techniques used 2-3	0.0225	0.1888	0.6639
Mother-baby relation Yes	-0.0417	0.7005	0.4026
GLOBAL		22.0317	0.1422

Table 4.11: Test for non-proportionality based on the scaled Schoenfeld residuals.

in Table 4.12 that these individuals' BF durations are high (censored at the end of the study) despite having a risk score slightly higher than one. The risk score is an estimate of the individuals' hazard ratio relative to the standard individual, calculated as the exponential of the linear predictor, using the covariates' values and the estimated coefficients. The omission of individuals 2 and 257, both separately and jointly, leads to a more negative parameter estimate for *education* and therefore to an increase in the significance of the effect (Table A.2 and Table A.3). In this model, the linearity test for the scaled Schoenfeld residuals does not lead to a rejection of the proportional hazard assumption at the 5% significance level (Figure 4.16 and Table A.4). Because of the outliers' influence and the actual shape of the non-parametric regression curve, the linearity test result is not valued and the original fitted model is considered adequate.

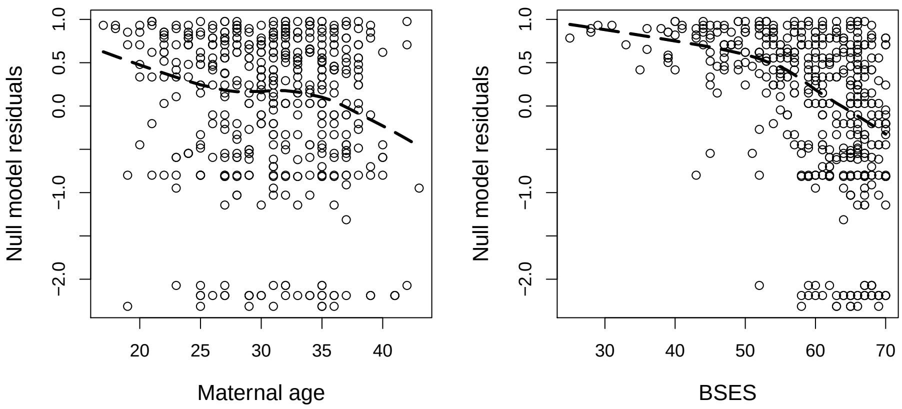


Figure 4.12: Martingale residuals for the null model against maternal age and bses score.

id	time	event	prognostic index	risk score	maternal age	BSES
2	19.2	0	0.046	1.048	35	65
257	17.8	0	0.271	1.312	42	52

Table 4.12: Predicted values, BF duration and covariates values for individuals 2 and 257.

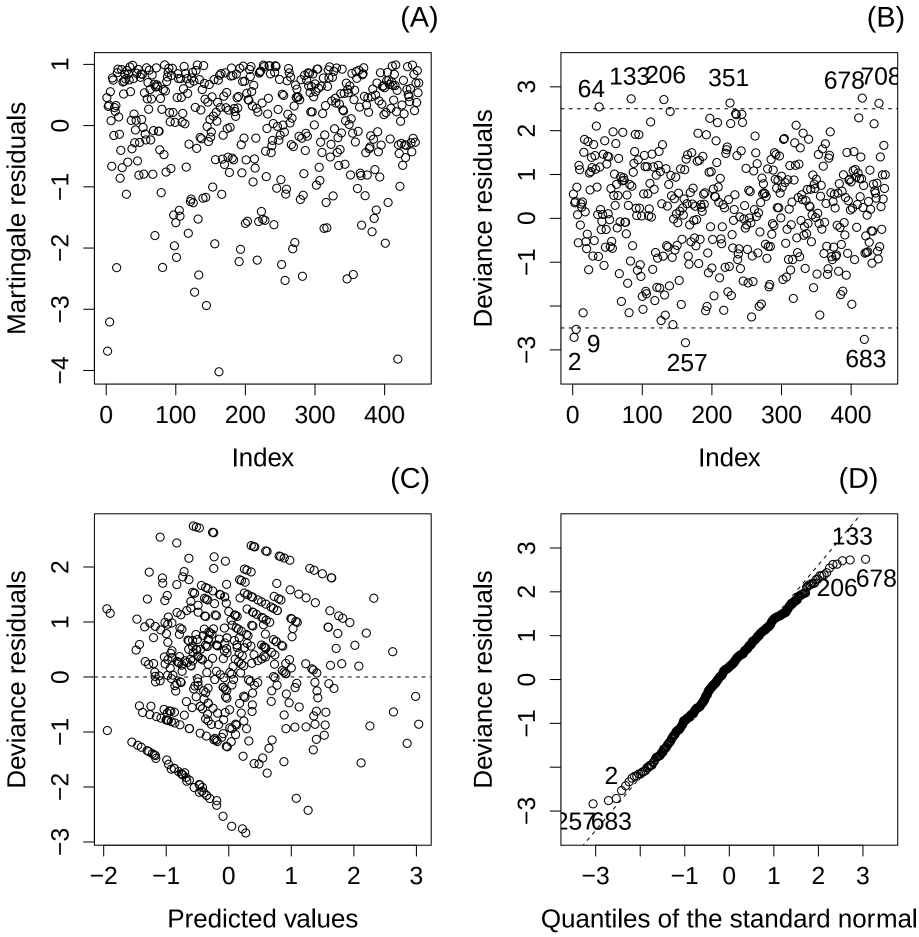


Figure 4.13: Martingale (A) and deviance (B) residuals against the individuals' index, deviance residuals against predicted values (C) and quantil-quantil plot of deviance residuals (D).

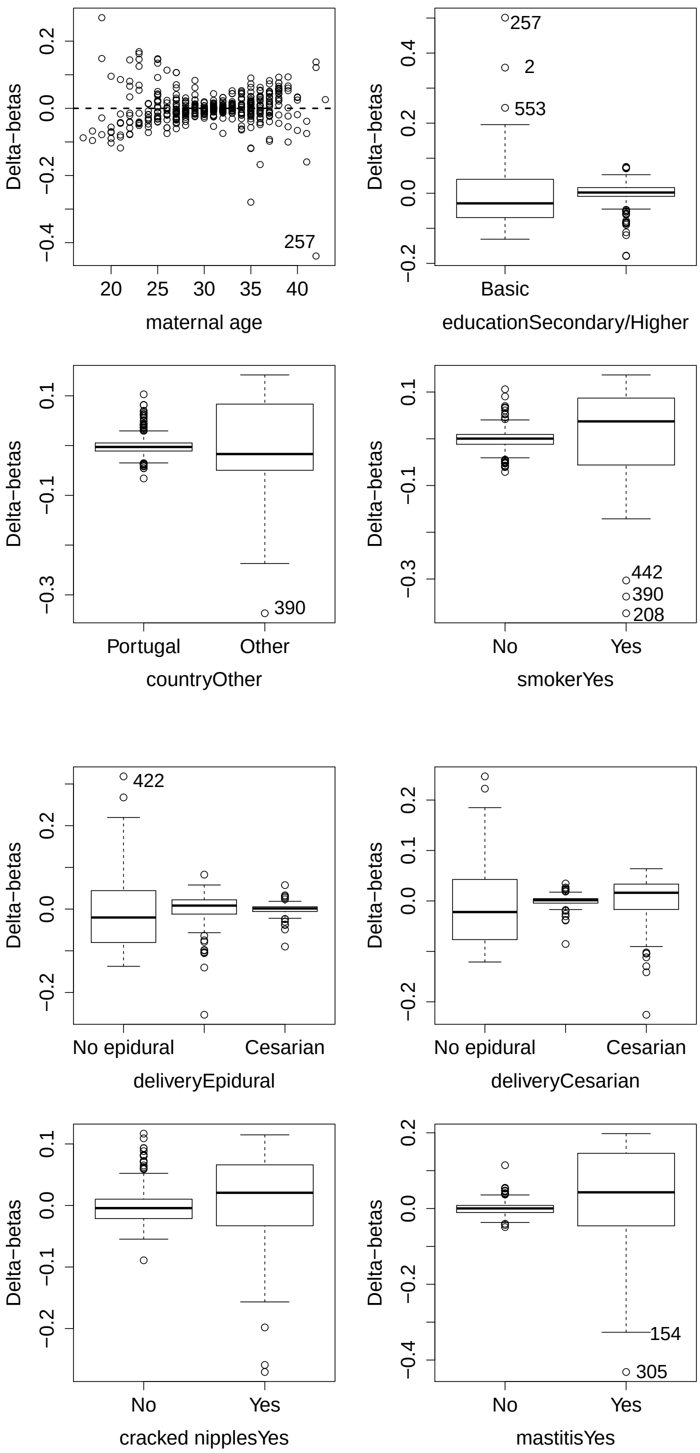


Figure 4.14: Standardized delta-betas against the values/categories of covariates. Individuals with absolute values above 0.3 are identified.

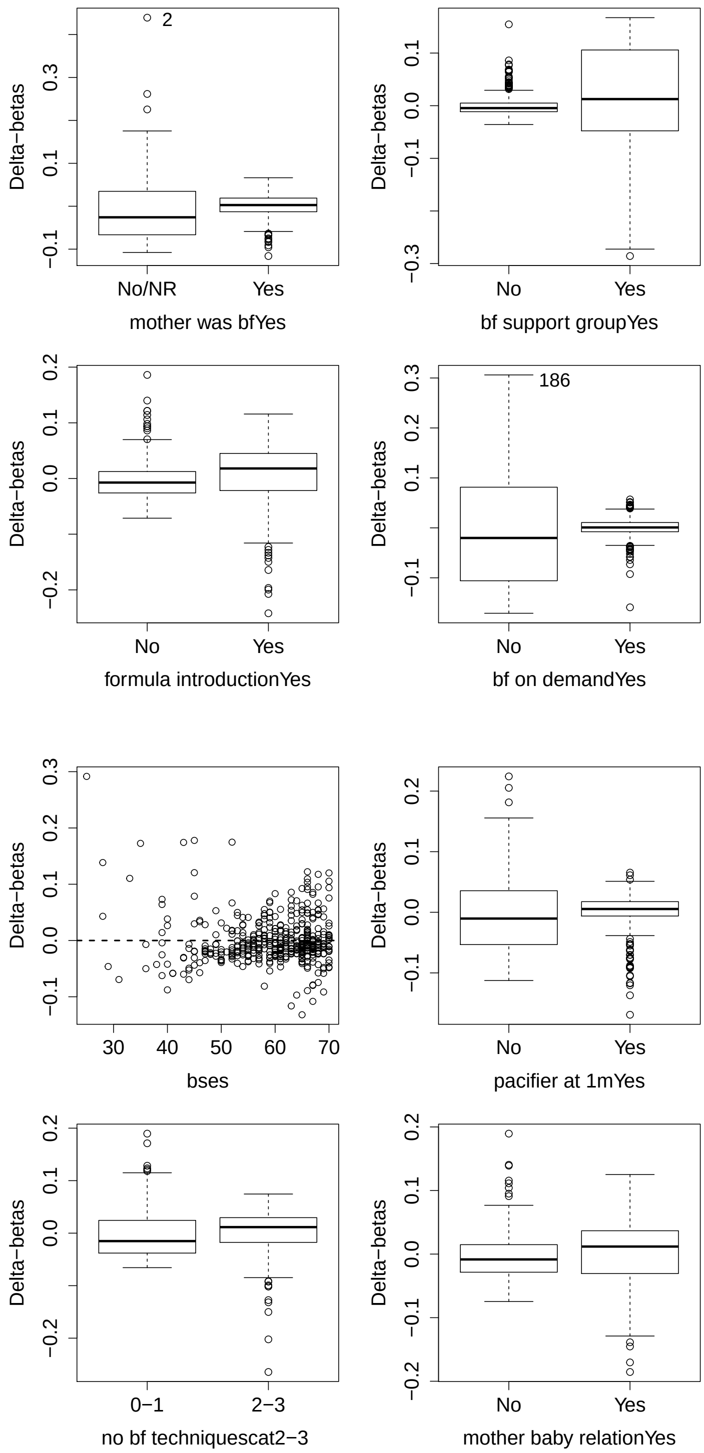


Figure 4.15: Standardized delta-betas against the values/categories of covariates. Individuals with absolute values above 0.3 are identified.

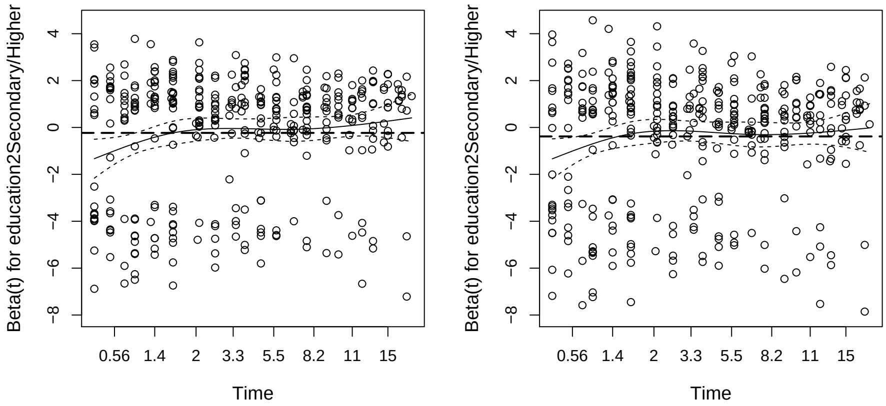


Figure 4.16: Scaled Schoenfeld residuals for variable *education* for the fitted model (left) and the same model estimated without observations 2 and 257.

4.3.2 Model adjustment

As shown in Table 4.13, the final model explains 36.3% of the variability in the BF duration, of which close to 21% are achieved with the four psychological variables included at the last level of modelling. The discriminatory power of the model is very good, with a concordance probability is 0.735. The high discriminatory power and good fit to the data can be visualized in Figure 4.17, where a good separation between the three groups is observed as well as an overlap between KM curves and those predicted by the fitted model (Carvalho *et al.*, 2011).

	Log likelihood	R^2	% variability explained
Null	-1949.97	0.000	0.0
Level 1	-1938.16	0.051	5.2
Level 2	-1927.67	0.095	9.5
Level 3	-1912.23	0.155	15.6
Model	-1850.01	0.359	36.3
Full	-895.08	0.991	100.0

Table 4.13: Log-likelihood, R^2 and percentage variability explained by the final model and at the end of each level.

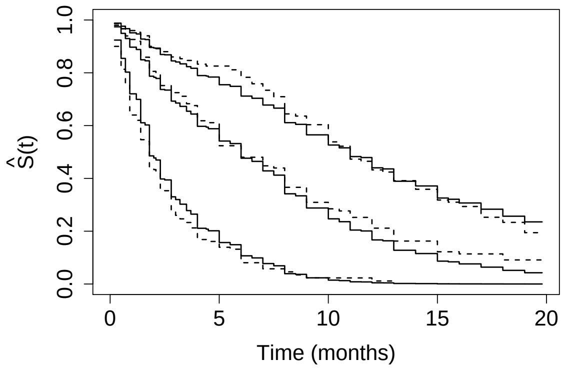


Figure 4.17: Survival curves, stratified by prognostic index, for the KM estimate (dashed line) and the fitted model (solid line).

4.4 Censored observations

The logistic model for censoring due to loss to follow up as a function of the covariates included in the final Cox model associated censored observations with being an immigrant and, to a much lower extent, with having a lower level of education and a higher BSES score (Table 4.14). Therefore, there is some evidence of censoring associated with covariates (informative censoring) which leads to loss of efficiency (Rocha & Papoila, 2009). However, it is assumed that censoring is independent, as defined in section 2.3, which is necessary for the validity of the conclusions taken.

4.5 Discussion and conclusion

This chapter presents the construction of a survival analysis model that identifies factors associated with the risk of cessation of any BF until eighteen months. The choice to model *any BF* instead of *exclusive BF* allowed the analysis of a higher number of individuals and for a longer period of time. Factors identified are in accordance with those reported in other studies.

Among demographic factors, maternal age, level of education and country of origin were retained in the model. A mother with a secondary or higher education has a risk of BF cessation 21% lower than one with a basic education level. Higher

	β_j	Pr(>|z|)
(Intercept)	-4.7351	0.0290
Maternal age	-0.0446	0.1610
Education Secondary/Higher	-0.6405	0.0967
Country Other	1.4567	0.0001
Smoked during pregnancy Yes	-0.8772	0.1774
Method of delivery Epidural	0.1764	0.6952
Method of delivery Cesarian	0.2452	0.6035
Cracked nipples Yes	0.6261	0.1099
Mastitis Yes	-0.0288	0.9710
Mother was BF Yes	0.1776	0.6753
Used support group Yes	-0.0690	0.8915
Formula introduction Yes	-0.5068	0.1934
BF on demand Yes	0.9865	0.2088
BSES score	0.0493	0.0593
Pacifier at age 1 month Yes	-0.4316	0.2252
No. BF techniques used 2-3	0.1607	0.6291
Mother-baby relation Yes	0.2159	0.5206

Table 4.14: Logistic regression output for censoring due to loss to follow up. Complete case analysis for n=455 individuals, including 48 censored observations.

maternal age and educational level are consistently associated in the literature with longer durations of BF (Forster *et al.*, 2006; Scott *et al.*, 2006).

Bivariate analysis revealed substantial differences between methods of delivery an epidural use, that do not remain significant in the final model. Normal delivery with epidural use was found to have a 24% increased risk of BF cessation compared to normal delivery without epidural. Most studies about BF determinants do not discriminate on the use of epidural, but cesarian delivery has been reported to have a more negative impact on BF than normal delivery (Al-Sahab *et al.*, 2010; Declercq *et al.*, 2009). However, the effect of the delivery method here described is supported by one study focusing specifically on intra-partum variables, in which the use of pharmacological methods of pain relief were significantly associated with an increased risk of BF cessation (Torvaldsen *et al.*, 2006). Another study found that the negative effect of intra-partum interventions was no longer significant after controlling for other confounders. There was, nonetheless, a cumulative effect of multiple interventions on the risk of BF cessation (Bai *et al.*, 2013).

Mothers who smoked during pregnancy or experienced major BF complications until three months have a higher risk of BF cessation. Mastitis was associated with an increase in risk of nearly 70%.

Early introduction of formula and pacifier use before one month of age were both strongly associated with a higher risk of BF cessation in the bivariate analysis. However, only pacifier use remained significant in the final model. It is associated with a 39% increase in risk of BF cessation. Pacifier use is claimed to reduce the frequency of BF episodes and consequently to shorten BF duration. In observational studies it is frequently associated with shorter durations of BF. Nonetheless, it is not certain whether this relation is causal or if pacifier use is a sign of other problems that are themselves the cause early weaning (Howard *et al.*, 2003; Kramer *et al.*, 2001).

The BSES score, a measure of confidence in the ability to breastfeed, remains highly significant in the final model ($p<0.0001$). Still, it is important to note that this variable was obtained at three months, when approximately 40% of the mothers had already terminated BF. It is possible that the highly significant association of this variable with the outcome is in part due to this fact.

Finally, it is worth to notice that both a higher number of BF techniques reported and the acknowledgement of the beneficial effect of BF on the mother-baby relationship remained significant the final model, both associated with a 20% decrease in risk.

Chapter 5

Logistic Regression

Logistic regression is a generalized linear model for a dependent variable of binary nature. The outcome is referred to generically as *success* and *failure*.

Let Y be this random variable that takes values 1 or 0 with probabilities $P(Y = 1) = \pi$ and $P(Y = 0) = 1 - \pi$, respectively. For n such random variables, their joint probability is

$$\prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}. \tag{5.1}$$

In classic linear regression, the expected value of Y given x, $E(Y_i|x)$ or π_i, is expressed as a linear equation in x. In a generalized linear model, this linear relation exists not with π_i directly, but with a transformation of the expected value, $g(\pi_i)$ (Hosmer & Lemeshow, 2000).

Logistic regression uses the *logit* transformation, defined as:

$$g(\pi_i) = logit(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right). \tag{5.2}$$

The expected value of Y, $P(Y_i = 1)$, ranges between 0 and 1, but the logit transformation may be continuous and range between $-\infty$ and $+\infty$, depending on x. From equation 5.2, π_i can be expressed as:

$$\pi_i = \frac{exp(\eta_i)}{1 + exp(\eta_i)}, \tag{5.3}$$

where η is the linear predictor, $\boldsymbol{\beta' x}$.

5.1 Model formulation and estimation

Model coefficients, β_j, have a simple relation with the *odds ratio* (OR), an effect measure that expresses how much more likely it is for the outcome to be present among those with $x = 1$ than those with $x = 0$. For a binary independent variable, that takes values 1 and 0, this relation can be expressed as follows:

$$\frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} = \frac{exp(\beta_0+\beta_1)}{exp(\beta_0)} = exp(\beta_1), \tag{5.4}$$

where $\pi(1)$ is $P(Y = 1|x = 1)$ and $\pi(0)$ is $P(Y = 1|x = 0)$.

The OR varies between 0 and $+\infty$, with 1 being the null value, when both x categories have equal chances for the outcome. If $\beta<0$, then the $OR<1$, which means that the event is less likely for $x = 1$ than $x = 0$. Consequently, when $\beta>0$, the $OR>1$ and the odds for the event are higher for $x = 1$ (Hosmer & Lemeshow, 2000).

Confidence intervals (CI) for the coefficients are obtained from Wald's statistic and the corresponding CI for the OR are obtained by exponentiating these values:

$$exp\big[\hat{\beta}_j \pm z_{1-\alpha/2} \times \widehat{SE}(\beta_j)\big]. \tag{5.5}$$

Parameter estimates are obtained via maximization of the log-likelihood function, which can be expressed as:

$$l\left(\boldsymbol{\pi};\boldsymbol{y}\right) = \prod_{i=1}^{N}\left[y_i \ln \pi_i + (n_i - y_i)\ln(1-\pi_i) + \ln\binom{n_i}{y_i}\right], \tag{5.6}$$

where N is the number of covariate patterns and n_i is the number of individuals for the ith pattern. For grouped data, the deviance statistic takes the form

$$D = 2\sum_{i=1}^{N}\left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right], \tag{5.7}$$

where $\hat{y}_i = n_i\hat{\pi}_i$.

If each observation has a different covariate pattern, which is likely in the presence of continuous covariates, then n_i=1 (Dobson, 2002).

Under the hypothesis that the model adjusts to data, the deviance statistic has an asymptotic chi-squared distribution with $n-p-1$ degrees of freedom, where p is the number of covariates. However, for non-grouped data, the approximation may be poor (Dobson, 2002; Hosmer & Lemeshow, 2000).

5.2 Model diagnostics

5.2.1 Residual analysis

Pearson and deviance residuals

For the binomial model, the standardized Pearson residual is given by

$$r_i^{p'} = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)(1-h_{ii})}}, \tag{5.8}$$

where h_{ii} is the leverage or hat-value of the ith observation. The leverage is a measure of the influence of y_i on $\hat{\pi}_i$. Therefore, the standardized Pearson residuals corrects for both the non-constant variance and the leverage of the observations (Fox & Weisberg, 2010).

The standardized deviance residuals are given by

$$r_i^{d'} = \frac{sgn(y_i - \hat{\pi}_i)\sqrt{d_i}}{\sqrt{1-h_{ii}}}, \tag{5.9}$$

where d_i is the individual component of the deviance statistic in equation (5.7).

The plots of Pearson and deviance residuals will be approximately normally distributed for grouped data with large n. However, for binary data or grouped data with small n the same cannot be expected (Dobson, 2002).

Influential observations

This measure allows the detection of influential observations by taking into account both leverage and Pearson residuals. Cooks distance is given by

$$D_i = \frac{r_i^p}{p} \times \frac{h_i}{(1-h_i)^2}, \tag{5.10}$$

where p is the number of parameters in the model.

Another useful diagnostic plot for influential observations is the plot of standardized residuals against hat-values.

Linear relation of covariates

The linear relation between a continuous covariate and the logit can be assessed through the plot of standardized residuals against the observed values of the variable.

Component plus residual plots are also useful to detect non-linearity (Fox & Weisberg, 2010). They display the partial relation between the outcome and a covariate, adjusting for all other covariates in the model. The partial residual of covariate x_j is given by

$$r_i^{(j)} = r_i + \hat{\beta}_j x_j, \tag{5.11}$$

where r_i is the ordinary residual $(y_i - \hat{\pi}_i)$. The plot of partial residuals against the values of x_j should reveal a linear relation, confirmed by a straight line for the smoothed curve.

5.2.2 Model Adjustment

For grouped data with large n, the deviance statistic can be used to assess model adjustment. For individual data, deviance will not be a useful measure of fit and a possible alternative is the Hosmer-Lemershow statistic (Dobson, 2002; Hosmer & Lemeshow, 2000).

The Hosmer-Lemershow statistic requires the creation of a number of groups based on predicted probabilities. The expected number of successes can be calculated for each group and compared to the observed values using Pearson's chi-squared statistic. For large samples,

$$X^2_{HL} = \sum_{k=1}^{g} \frac{(o_k - e_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \sim \chi^2_{g-2} \tag{5.12}$$

where g is the number of groups, o_k and e_k are respectively the observed and expected number of events for the kth group, n_k is the number of individuals in the group and $\bar{\pi}_k$ is the predicted risk for the group. Ten groups are commonly used.

Although not a measure of goodness of fit, a classification table is a useful way to summarise the results of a logistic model (Hosmer & Lemeshow, 2000). In this table, the observed response is cross-classified with a predicted outcome, based on the predicted probabilities. To obtain the predicted outcome it is necessary to define a cut-off value that discriminates successes and failures. The most commonly used cut-off value is 0.5.

A more complete measure of discriminatory ability is the area under the ROC (AUC) curve because it covers the entire range of possible cut-off values (Hosmer & Lemeshow, 2000). The AUC ranges between 0 and 1 and a higher value corresponds to a higher discriminatory ability.

Chapter 6

Application - Logistic Regression

The response variable *goal achievement* is a binary variable that defines BF success or failure by the achievement of an individual goal for BF duration.

This chapter describes the building of a model that aims to evaluate the influence of BF counselling by health professionals on the achievement of the goal. In particular, the exposure variables chosen refer to BF counselling about benefits, techniques and difficulties.

6.1 Response variable

A quantitative goal for BF duration was specified for 452 (95.5%) mothers. Its distribution is shown in Figure 6.1. A high proportion of mothers specify six (23.2%) and 12 (45.4%) months as planned duration. The median is 12 months.

Among these 452 mothers, 257 (56.9%) expected to BF until the milk ended and 90 (19.9%) planned to do so for as long as the baby wanted.

Table 6.1 shows a cross-table of *goal achievement* with the censoring indicator variable. Censored BF durations are responsible for the loss of 41 individuals, for whom the outcome is undetermined.

Table 6.2 shows a cross-table of the outcome variable with censoring at the three contact periods, for the 92 censored observations. Censoring at three and six months lead almost always to an undetermined outcome while censored observations at eighteen months were mostly successes.

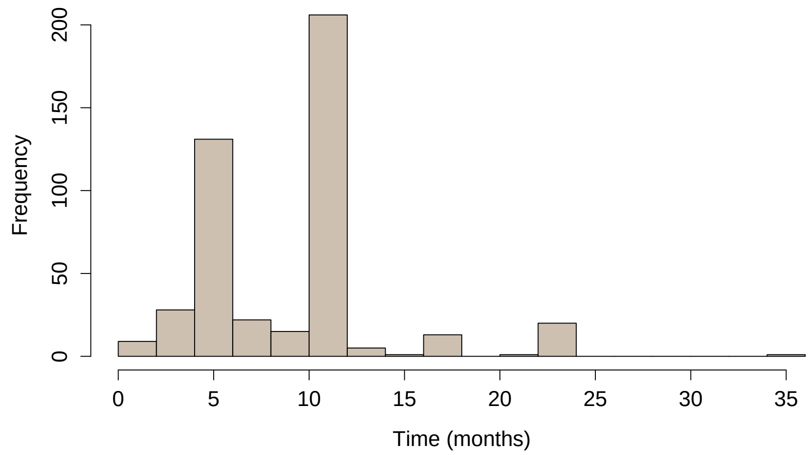


Figure 6.1: Histogram of planned BF duration (n=452).

In Figure 6.2, the three initial categories of the outcome variable are plotted against the quantiles of event time. It can be observed that the proportion of successes increases with time and the proportion of failures decreases with time. The proportion of *undetermined* outcome is higher at the second and third quantiles, which include censored observations at three and six months.

In Figure 6.3, the three-category outcome is plotted against categorized planned duration. The proportion of successes appears to be higher for smaller than for higher planned durations. It appears that smaller goals are easier to accomplish but the contrary could also be expected, as smaller planned durations could reflect lower motivations.

After the exclusion of individuals with *undetermined* outcome, the analysis proceeds with 411 cases, among which 142 (34.5%) achieved their goal and 269 (65.5%) failed to.

6.2 Variables of primary interest

Two variables that reflect BF counselling by healthcare professionals are chosen as being of primary interest in this model. In an epidemiology context these are referred as exposure variables or risk factors. They correspond to the questions

	Event		
Outcome	0	1	sum
achieved	44	98	142
failed	2	267	269
undetermined	46	16	62
sum	92	381	473

Table 6.1: Cross-table of *goal achievement* with censoring indicator (1 for observed and 0 for censored BF duration).

	Time of censoring			
Outcome	3M	6M	18M	sum
achieved	3	6	35	44
failed	2	0	0	2
undetermined	16	23	7	46
sum	21	29	42	92

Table 6.2: Cross-table of *goal achievement* with censoring at 3, 6 and 18 months, for the 92 censored observations.

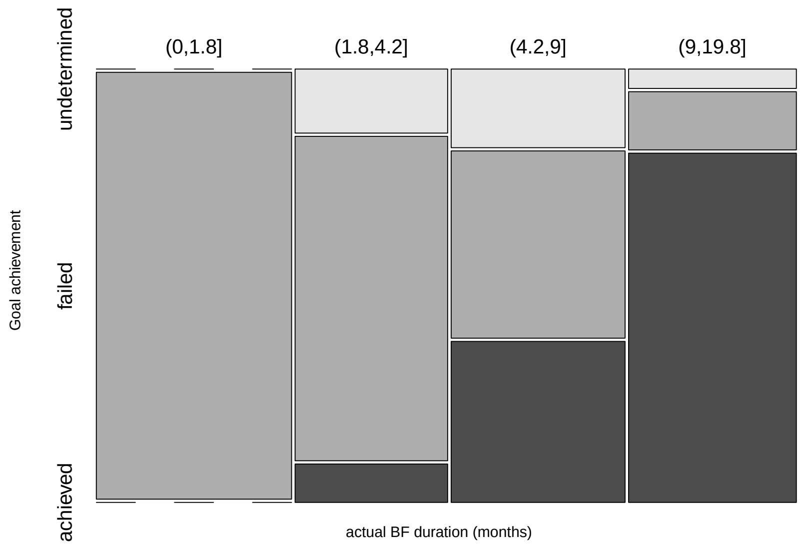


Figure 6.2: Proportion barplot of *goal achievement* by the quantiles of *BF duration* (n=452).*Goal achievement* categories are *achieved* (dark grey), *failed* (grey) and *undetermined* (light grey).

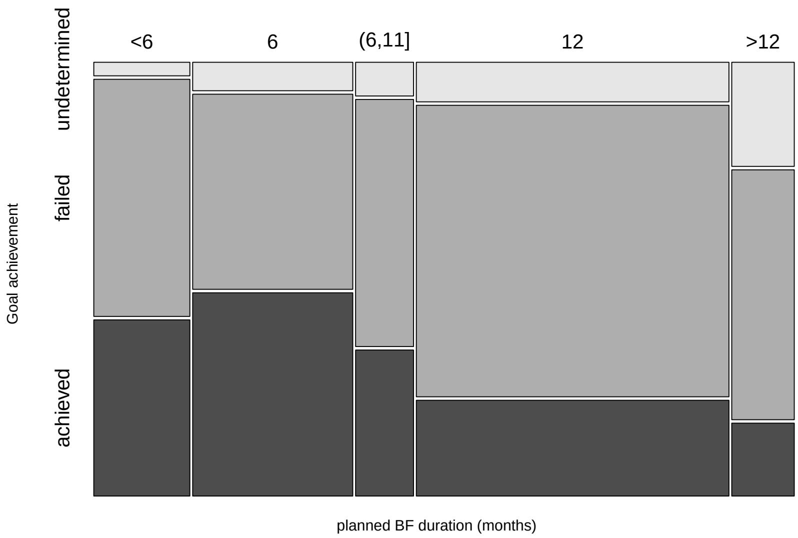


Figure 6.3: Proportion barplot of *goal achievement* by categorized *planned BF duration* (n=452).*Goal achievement* categories are *achieved* (dark grey), *failed* (grey) and *undetermined* (light grey).

where mothers were asked if they were ever told about BF benefits and techniques or difficulties. If so they were also asked when: prenatal consults, prenatal classes, at maternity or at postnatal consults. Both variables were categorized in three groups: was never counselled (*Never*), was first counselled before delivery (*Antenatal*) and was only counselled after delivery (*Postnatal*). Their distribution and unadjusted OR are shown in Table 6.3.

The majority of mothers recall having been counselled at least once - 77% for benefits and 71% for techniques/difficulties - while only 60 (14.5%) answer no to both questions. Mothers that recall having been counselled are slightly older (t-test p= 0.088 with mean age of 31.1 for counselled about benefits versus 30.0 for not counselled; no difference for techniques), more educated (chi-squared test of independence with p=0.117 for benefits and p<0.001 for techniques) and are having their first BF experience (p=0.022 for benefits and p=0.004 for techniques). Immigrants were more likely than Portuguese nationals to not have been counselled about techniques (p=0.160) but no differences were found for counselling

Variable	N	OR	95% CI	Wald's test p-value
Taught BF benefits				
Never	94			
Antenatal	212	1.22	0.73 - 2.06	0.4562
Postnatal	105	1.07	0.72 - 1.92	0.8313
Taught BF techniques or difficulties				
Never	118			
Antenatal	154	0.93	0.56 - 1.55	0.7784
Postnatal	139	1.06	0.63 - 1.77	0.8378

Table 6.3: Unadjusted associations between variables associated with the transmission of knowledge about BF and *goal achievement*.

about benefits (p=0.745).

6.3 Variable selection and model building strategy

Demographic and biological variables with an unadjusted OR significantly different from 1 at a level of 0.20 are reported in Table 6.4. However, because the model aims to evaluate two variables of primary interest, variable selection was not based solely on statistical significance.

The variables *maternal age*, *education*, *previous BF experience* and *country of origin* were considered as *a priori* confounders and were included independently of statistical significance. These variables were also investigated as effect modifiers, which must be biologically and statistically significant (Greenland, 1989).

Because the variables of interest belong to the social level of BF determinants, it is important to adjust for potential confounders from the higher levels, namely demographic and biological variables. No other social variables were considered for entry besides the two variables of interest. Variables influenced by or that can be viewed as an outcome of the exposure variables should not be included in the model (Greenland, 1989). An example of such variables are the psychological variables associated with maternal knowledge and skills.

The other independent variables considered were included in the model if considered as potential confounders and/or predictors of outcome. The following criteria was considered: effect modifiers with $p < 0.05$, confounders that change the OR of the exposure variables by 10% and predictors of outcome with $p < 0.20$.

Variable	OR	95% CI	p-value $H_0 : \beta_j$=0
Demographic			
Maternal age	1.09	1.04 - 1.14	0.0001
Marital status			
No partner			
Has a partner	1.93	1.03 - 3.86	0.0486
Country of origin			
Portugal			
Other	2.60	1.49 - 4.57	0.0008
Education			
Basic			
Secondary	1.55	0.87 - 2.83	0.1435
Higher	1.99	1.16 - 3.49	0.0135
Biological			
Previous BF experience			
No			
Yes	1.60	1.06 - 2.41	0.0250
Smoked during pregnancy			
No			
Yes	0.42	0.19 - 0.84	0.0192
Method of delivery			
Normal, no epidural			
Normal, epidural	0.65	0.36 - 1.15	0.1346
Cesarean	0.90	0.50 - 1.64	0.7249
Major complications [1]			
No			
Yes	0.62	0.39 - 0.97	0.0374
Postpartum depression (first 6M)			
No			
Yes	0.24	0.06 - 0.70	0.0210

Table 6.4: Unadjusted associations between demographic and biological variables with *goal achievement*.

[1]Complications: blocked ducts, mastitis and cracked nipples.

6.4 Model Building

The sequence of adjusted models, including estimated coefficients and OR, is shown in Table 6.5.

Maternal age, *education*, *previous BF experience* and *country of origin* are included in a prespecified order. The order of remaining biological variables, *method of delivery* and *complications*, is determined by the deviance significance.

Maternal age is determinant for the outcome, leading to a change in deviance with p=0.0001. *Education* is included with a change in deviance with p=0.1638. Both variables are potential confounders, causing changes in the estimated effects for *benefits postnatal* and *techniques prenatal*. *Previous BF experience* did not cause a significant change in deviance or change in the estimated effects. The inclusion of *country of origin* was significant (p=0.0159) and lead to an 11% increase in the coefficient of *techniques postnatal*. *Complications* were also significantly determinant to explain the outcome (p=0.0191). *Method of delivery* did not lead to a significant change in deviance or OR (p=0.6082) and was not included in the final model (model 3 in Table 6.5).

Interaction terms between *a priori* confounders and exposure variables were all above the 5% significance level. However, it is worth to note that a borderline significant interaction between *country of origin* and *techniques* was detected (p=0.0742). This model is shown in Table B.1.

	(Model 1)			(Model 2)			(Model 3)		
	β_j	$exp(\beta_j)$	p-value $H_0 : \beta_j$=0	β_j	$exp(\beta_j)$	p-value $H_0 : \beta_j$=0	β_j	$exp(\beta_j)$	p-value $H_0 : \beta_j$=0
(Intercept)	-3.346	0.04	$<$0.0001	-3.031	0.05	0.0001	-2.862	0.06	0.0003
Counselled about benefits									
Never									
Antenatal	0.398	1.49	0.2229	0.392	1.48	0.2368	0.481	1.62	0.1534
Postnatal	-0.104	0.90	0.7551	-0.069	0.93	0.8385	-0.063	0.94	0.8531
Counselled about techniques/difficulties									
Never									
Antenatal	-0.494	0.61	0.1439	-0.327	0.72	0.3443	-0.336	0.72	0.3348
Postnatal	-0.015	0.99	0.9593	0.019	1.02	0.9476	0.058	1.06	0.8458
Maternal age	0.074	1.08	0.0013	0.056	1.058	0.0314	0.052	1.05	0.0477
Level of education									
Basic									
Secondary	0.447	1.56	0.1536	0.342	1.41	0.2863	0.414	1.51	0.2015
Higher	0.559	1.75	0.0640	0.526	1.69	0.0873	0.609	1.84	0.0507
Previous BF experience									
No									
Yes				0.240	1.27	0.3549	0.250	1.28	0.3408
Country of origin									
Portugal									
Other				0.723	2.06	0.0159	0.724	2.06	0.0162
Major complications									
No									
Yes							-0.569	0.57	0.0213

Table 6.5: Coefficients and OR for model 1 (adjusted for age and education), model 2 (further adjusted for previous BF experience and country of origin) and model 3 (further adjusted for complications). Model 3 is the final model at this stage (variable selection). All models were fitted to n=410 individuals.

Model	Resid. Df	Resid. Dev	AIC	Df	Deviance	Pr(>Chi)	
(1) linear	399	495.48	517.48				
(2) degree 2 polynomial	398	487.40	511.40	1	8.0765	0.0045	(2) vs (1)
(3) degree 3 polynomial	397	482.22	508.22	1	5.1755	0.0229	(3) vs (2)
(4) cubic spline (df=3)	397	483.63	509.63				

Table 6.6: Deviance and AIC for models with different transformations to model the relation between *maternal age* and the logit. See also Appendix C

6.5 Model Diagnostics

6.5.1 Linearity assessment

While observing the plots of deviance residuals for the fitted model (model 3) it stands out that *maternal age* might have a non-linear relation with the response (Figure 6.4 A). This is confirmed in the component-plus-residual plot in Figure 6.4 B and by the change in deviance caused by the inclusion of a quadratic term (p=0.0045).

The relation between *maternal age* and the logit was modelled through the inclusion of higher order terms (polynomial regression) and natural cubic splines (function *ns* from library *splines*). The models fitted are shown in Table 6.6.

The inclusion of quadratic and cubic terms leads to a significant change in deviance (p=0.0045 and p=0.0229, respectively). The model with the cubic spline shown in Table 6.6 has 3 coefficients and 2 internal knots. Splines with one, two and three internal knots were tested and the choice between them was based on *Akaike's information criterion* (AIC), according to Harrell (2001). *Effect displays* for the models in Table 6.6 are shown in Figure 6.5. The plot for the cubic spline effect is similar to the cubic polynomial. Both plots show a linear increase in the fitted values of *goal achievement* up to 30 years of age. Beyond this point, an increase in age doesn't change the fitted values. Both models have lower AIC than the linear model and the quadratic polynomial (508.2 for the cubic polynomial and 509.6 for the cubic spline).

According to Harrell (2001), regression splines have several advantages over polynomials. Therefore, the cubic spline model was chosen to correct for the non-linearity in *maternal age*. It is shown in Table 6.7.

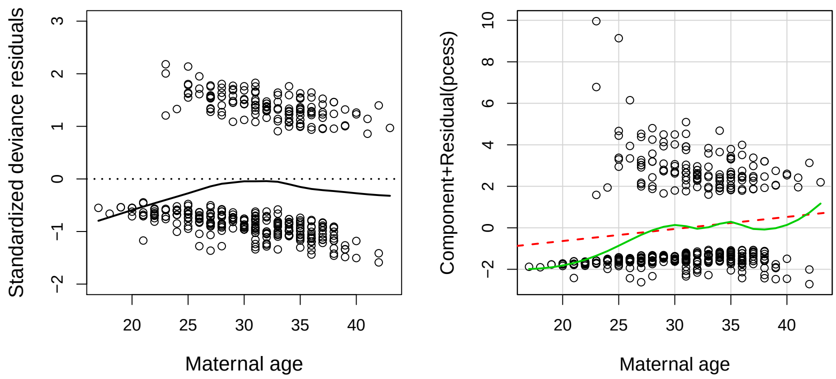


Figure 6.4: Standardized deviance residuals and component plus residual plot for *maternal age*.

	β_j	SE	p-value $H_0 : \beta_j{=}0$	$exp(\beta_j)$	95% CI
(Intercept)	-5.361	1.355	0.0001		
Counselled about benefits					
Never					
Antenatal	0.541	0.345	0.1165	1.72	0.88 - 3.41
Postnatal	-0.102	0.344	0.7660	0.90	0.46 - 1.78
Counselled about techniques/difficulties					
Never					
Antenatal	-0.464	0.357	0.1929	0.63	0.31 - 1.26
Postnatal	0.099	0.301	0.7429	1.10	0.61 - 2.00
Maternal age					
spline term 1	2.273	0.760	0.0028		
spline term 2	8.977	2.941	0.0023		
spline term 3	1.723	0.935	0.0654		
Level of education					
Basic					
Secondary	0.228	0.333	0.4947	1.26	0.66 - 2.43
Higher	0.387	0.319	0.2251	1.47	0.79 - 2.78
Previous BF experience					
No					
Yes	0.297	0.268	0.2685	1.35	0.80 - 2.28
Country of origin					
Portugal					
Other	0.778	0.304	0.0106	2.18	1.20 - 3.98
Major complications					
No					
Yes	-0.597	0.249	0.0166	0.55	0.33 - 0.89

Table 6.7: Model with cubic spline in *maternal age* (n=410).

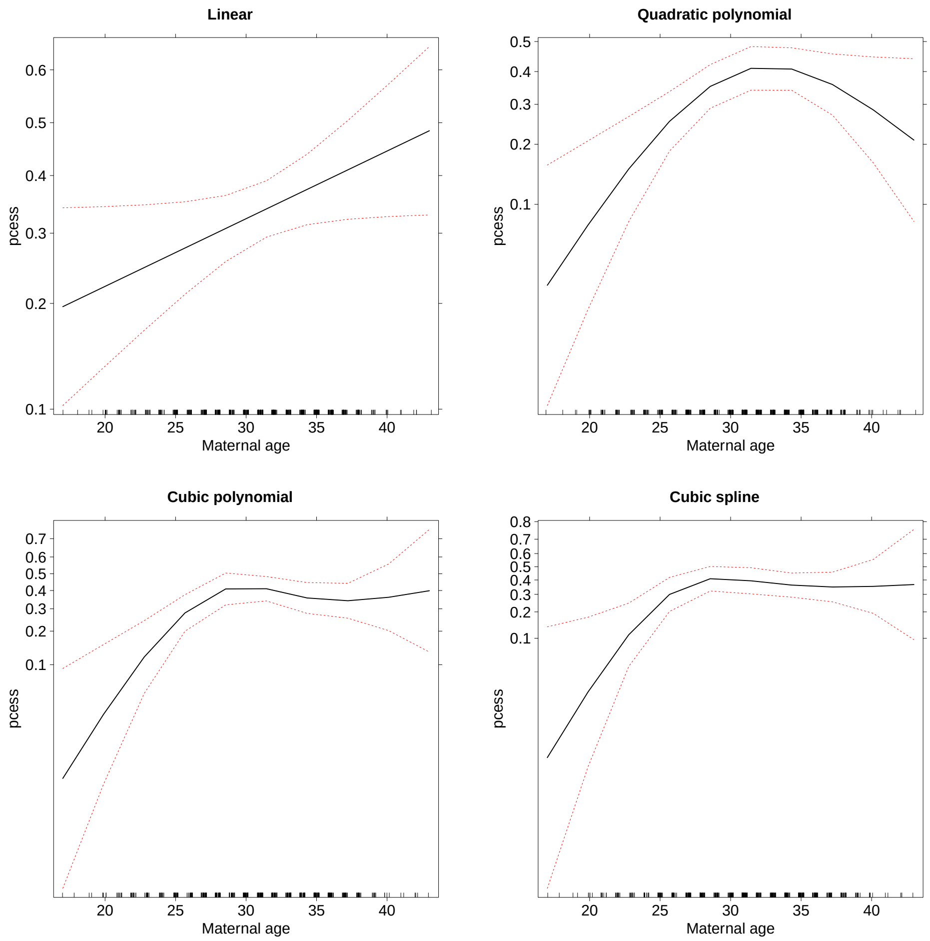


Figure 6.5: *Effect displays* for the models in Table 6.6, with a linear term, degree 2 and 3 polynomials and a cubic natural spline for the relation between *maternal age* and the logit.

6.5.2 Residual analysis

Standardized deviance residuals for the adjusted model in Table 6.7 are shown in Figure 6.6. As expected, because the data is bernoulli, the residuals do not approximate the normal distribution. They have an asymmetric distribution, with most values between -2 and 2. The smoothed line of the residuals against *maternal age* confirm the non-linearity was corrected.

Figure 6.7 shows plots of leverage and Cook's distance. Although there are a number of observations with high leverage, none exhibited simultaneously high

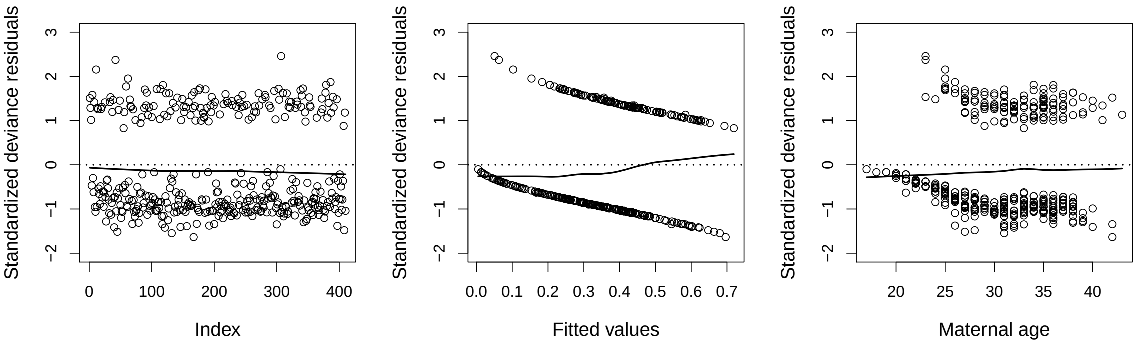


Figure 6.6: Standardized deviance residuals against index, fitted values and *maternal age*.

leverage and residual values.

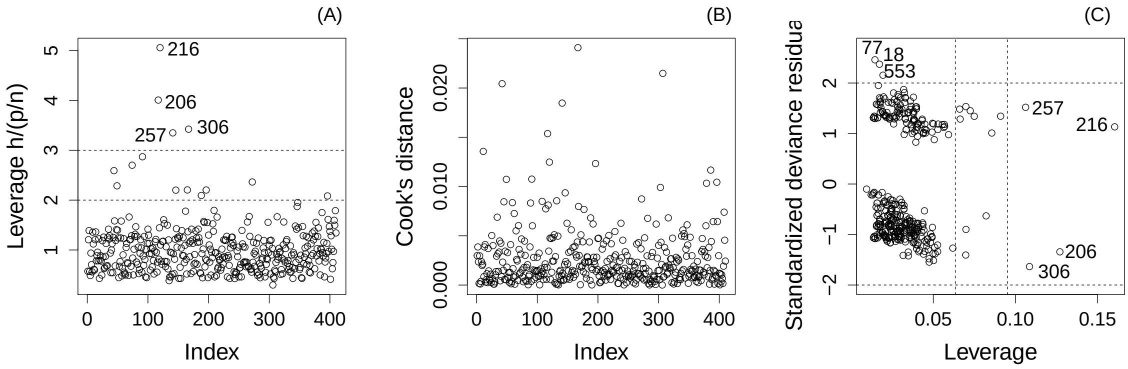


Figure 6.7: Plots to detect influential observations: values of leverage (A) and Cook's distance (B) and deviance residuals against leverage (C).

6.5.3 Model adjustment

The result for the Hosmer-Lemershow test shown in Table 6.8 indicates that the model fits the data well (p=0.6362).

Although prediction is not the primary goal of the model, measures based on predictive ability are useful to understand how the model performs. The percentage of correct predictions is 69,0%, 36.5% for successes and 91.4% for

		Achieved (y=1)		Failed (y=0)	
Decile	Prob.	Obs	Exp	Obs	Exp
1	0.125	3	2.6	38	38.4
2	0.231	5	7.6	36	33.4
3	0.273	12	10.3	29	30.7
4	0.320	13	12.0	28	29.0
5	0.357	13	14.0	28	27.0
6	0.400	17	15.5	24	25.5
7	0.427	12	17.0	29	24.0
8	0.455	22	17.9	19	23.1
9	0.546	21	20.8	21	21.2
10	0.719	24	24.3	16	15.7
		χ_{HL} = 6.099		p=0.6362	

Table 6.8: Hosmer-Lemershow test: observed and expected frequencies for each of the ten deciles of risk.

	Predicted group		
Actual group	failed	achieved	Percent correct
failed	245	23	91.4
achieved	104	38	36.5
Percent of cases correctly classified			69.0

Table 6.9: Classification table for the model using a cut-off value of 0.5.

failures, as shown in Table 6.9.

Area under ROC curve is 69.2%, which according to Hosmer & Lemeshow (2000) is close to acceptable discriminatory ability. It means that for 69.2% of pairs with different observed outcomes (one success and one failure) the model correctly assigned a higher probability to the success.

6.6 Discussion and conclusion

With this analysis, an attempt is made to explain the role of BF counselling by healthcare professionals in BF success, defined as the achievement of the intended BF duration. The variables of interest expressed if the mother was ever counselled about benefits or techniques/difficulties and if so, the timing when it first occurred. Although there was at least another question in the survey that was more informative about counselling in the postnatal period, the question was only applied to very few women that made it impossible to use for an inferential goal.

The final model is presented in Table 6.7. Among the variables of interest, only counselling about BF benefits exhibited a borderline association with BF success. Prenatal counselling about benefits was associated with a 70% increase in the odds of success, compared to not having been counselled (p=0.1165). No association was found for counselling about benefits in the postnatal period, as well as counselling about BF techniques or difficulties in both periods. Maternal age and nationality remain significantly associated with BF success in the final model. The relation between maternal age and BF success is not linear and had to be modelled through a spline function. Maternal age was found to have a linear positive relation with BF success until approximately 30 years of age, after which any further increase does not have an effect. Being an immigrant doubles the odds of success. Education and having had a previous BF experience are not significant in the final model. Having had major BF complications, such as blocked ducts, mastitis and cracked nipples, was associated with a 45% reduction in the odds of success.

Interventional studies have found a significant effect of BF education and healthcare support, at both prenatal and postnatal periods, on success (at specified time points) (Mattar *et al.*, 2007; Su *et al.*, 2007). On the contrary, data from this study is purely observational. Not only it relies on the memory and acknowledgement of having been counselled, but it was also found to be vulnerable to confounding. A number of these confounders were adjusted for. However, other important confounders may have been missed.

It is also possible that awareness of BF benefits has an effect on both intended and actual duration. However, higher intended durations may be more difficult to achieve. If this is the case, then a possible positive effect would be antagonized.

It is important to note that although the proportion of successes was approximately one third among the 411 women included, there were 49 women excluded that were censored before their intended duration. As such, there is some differential loss: women who breastfed longer are more likely to be excluded because of loss to follow up. Therefore, it is likely that the actual proportion of women achieving their BF goal is higher.

Chapter 7

Conclusion

This work intends to contribute to the understanding of maternal BF in Portugal, in particular through the identification of potentially modifiable determinants.

The first part of this thesis included a proportional hazards model for BF duration in order to identify important predictors. For modelling, a hierarchy was established between demographic, biological, social and psychological variables. In this manner, an attempt is made to account for the temporal and causal relations among the different groups and to adjust for potential confounding.

This model identified known demographic factors, such as maternal age and education, but also modifiable factors associated with the health care system and psychological maternal factors, such as early introduction of formula and pacifier and the knowledge of BF techniques. Psychological factors accounted for an important proportion of the model's explained variability.

For the second part of this work, a binary outcome for BF was defined as the achievement of an individually specified duration. A logistic model was then applied in order to investigate the potential influence of BF counselling by health-care professionals on BF success. In particular, the variables of interest referred to BF counselling about benefits, techniques and difficulties, and the its timing.

Only counselling about BF benefits exhibited a borderline association with BF success. Prenatal counselling about benefits was associated with a 70% increase in the odds of success, compared to not having been counselled. An effect for counselling about BF techniques or difficulties was not found.

There are several limitations to the conclusions that can be drawn from this model. The data is observational and is, therefore, subject to substantial confounding. Also, the questions applied were too general and may include different interventions by healthcare professionals, which may have different effects. Finally, there was some differential loss with individuals who breastfed longer more likely to be excluded because of loss to follow up.

Appendix A

Cox Regression - Supplementary material

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
maternal_age	-0.0200	0.9802	0.0112	-1.7849	0.0743
education2Secondary/Higher	-0.2323	0.7927	0.1342	-1.7309	0.0835
countryOther	-0.3851	0.6804	0.1640	-2.3478	0.0189
active_smokerYes	0.3491	1.4177	0.1637	2.1328	0.0329
deliveryEpidural	0.2134	1.2379	0.1569	1.3603	0.1737
deliveryCesarian	0.1123	1.1189	0.1633	0.6880	0.4915
cracked_nipplesYes	0.0824	1.0859	0.1345	0.6128	0.5400
mastitisYes	0.5283	1.6960	0.2243	2.3552	0.0185
mother_was_bfYes	-0.3186	0.7272	0.1346	-2.3664	0.0180
hc_usedYes	-0.2802	0.7556	0.1776	-1.5776	0.1147
formulaintYes	0.0583	1.0601	0.1157	0.5042	0.6141
bfondemandYes	-0.3892	0.6776	0.1855	-2.0979	0.0359
bses	-0.0722	0.9303	0.0068	-10.6527	0.0000
pacifier1mYes	0.3299	1.3909	0.1335	2.4708	0.0135
nopractices_cat2-3	-0.2294	0.7950	0.1081	-2.1218	0.0339
mbrelationshipYes	-0.2128	0.8083	0.1141	-1.8642	0.0623

Table A.1: Summary output for the fitted model.

	$\hat{\beta}_j$	$\hat{\beta}_{j(-2)}$	$\hat{\beta}_{j(-257)}$	$\hat{\beta}_{j(-c(2,257))}$
maternal_age	-0.0200	-0.0162	-0.0138	-0.0092
education2Secondary/Higher	-0.2323	-0.2902	-0.3154	-0.3832
countryOther	-0.3851	-0.3949	-0.3938	-0.4056
active_smokerYes	0.3491	0.3417	0.3469	0.3393
deliveryEpidural	0.2134	0.2614	0.2314	0.2848
deliveryCesarian	0.1123	0.1291	0.1584	0.1806
cracked_nipplesYes	0.0824	0.0969	0.0632	0.0776
mastitisYes	0.5283	0.5420	0.5191	0.5330
mother_was_bfYes	-0.3186	-0.3898	-0.2995	-0.3765
hc_usedYes	-0.2802	-0.3139	-0.2975	-0.3367
formulaintYes	0.0583	0.0411	0.0880	0.0716
bfondemandYes	-0.3892	-0.3818	-0.3901	-0.3825
bses	-0.0722	-0.0712	-0.0737	-0.0727
pacifier1mYes	0.3299	0.3428	0.3453	0.3610
nopractices_cat2-3	-0.2294	-0.1952	-0.2023	-0.1628
mbrelationshipYes	-0.2128	-0.2324	-0.1866	-0.2057

Table A.2: Effect of the removal of individuals 2 and 257 on parameter estimates.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
maternal_age	-0.0092	0.9908	0.0116	-0.7947	0.4268
education2Secondary/Higher	-0.3832	0.6817	0.1380	-2.7765	0.0055
countryOther	-0.4056	0.6666	0.1639	-2.4746	0.0133
active_smokerYes	0.3393	1.4039	0.1633	2.0776	0.0377
deliveryEpidural	0.2848	1.3295	0.1580	1.8028	0.0714
deliveryCesarian	0.1806	1.1979	0.1639	1.1014	0.2707
cracked_nipplesYes	0.0776	1.0807	0.1347	0.5765	0.5643
mastitisYes	0.5330	1.7041	0.2246	2.3735	0.0176
mother_was_bfYes	-0.3765	0.6863	0.1365	-2.7577	0.0058
hc_usedYes	-0.3367	0.7141	0.1785	-1.8858	0.0593
formulaintYes	0.0716	1.0743	0.1154	0.6207	0.5348
bfondemandYes	-0.3825	0.6822	0.1853	-2.0638	0.0390
bses	-0.0727	0.9299	0.0067	-10.7726	0.0000
pacifier1mYes	0.3610	1.4348	0.1335	2.7049	0.0068
nopractices_cat2-3	-0.1628	0.8498	0.1090	-1.4930	0.1354
mbrelationshipYes	-0.2057	0.8141	0.1142	-1.8013	0.0717

Table A.3: Summary output for the model without individuals 2 and 257.

	rho	chisq	p
maternal_age	0.0712	2.0732	0.1499
education2Secondary/Higher	0.0882	2.9324	0.0868
countryOther	0.0423	0.6718	0.4124
active_smokerYes	0.0164	0.1013	0.7502
deliveryEpidural	-0.0230	0.1980	0.6564
deliveryCesarian	-0.0494	0.9193	0.3377
cracked_nipplesYes	0.0728	2.2020	0.1378
mastitisYes	0.0383	0.5820	0.4455
mother_was_bfYes	-0.0918	3.4011	0.0652
hc_usedYes	-0.0206	0.1621	0.6872
formulaintYes	-0.0707	1.9816	0.1592
bfondemandYes	0.0313	0.3897	0.5325
bses	-0.0561	1.1423	0.2852
pacifier1mYes	-0.0574	1.2237	0.2686
nopractices_cat2-3	0.0483	0.8851	0.3468
mbrelationshipYes	-0.0486	0.9499	0.3297
GLOBAL		20.0782	0.2167

Table A.4: Effect of the removal of individuals 2 and 257 on the test for non-proportionality based on the scaled Schoenfeld residuals.

Appendix B

Logistic Regression - Supplementary material

	β_j	$exp(\beta_j)$	p-value $H_0 : \beta_j$=0
(Intercept)	-3.1046	0.0448	0.0001
benefitsAntenatal	0.4087	1.5049	0.2264
benefitsPostnatal	-0.1103	0.8956	0.7503
techniquesAntenatal	-0.0614	0.9405	0.8709
techniquesPostnatal	0.3204	1.3777	0.3370
maternal_age	0.0522	1.0536	0.0490
educationSecondary	0.3384	1.4027	0.2949
educationHigher	0.4728	1.6044	0.1298
previousbfexpYes	0.2636	1.3017	0.3151
countryOther	1.6884	5.4108	0.0020
techniquesAntenatal:countryOther	-1.4694	0.2301	0.0547
techniquesPostnatal:countryOther	-1.4476	0.2351	0.0536

Table B.1: Model with a borderline significant interaction term for *country:techniques*

Appendix C

R code

Data analysis was carried out using R (version 2.15.1) (R Core Team, 2012).

Figure 4.17 was made using function *plot.pi()* referenced in Carvalho *et al.* (2011) and accessed online ⟨`http://sobrevida.fiocruz.br/rlatex.html`⟩.

Variable selection using the stepwise method described in Section 4.2 was carried out using an algorithm adapted from Leite (2012). The excerpt shown was used in level 1, step 1 of Cox regression.

```r

deviance <- function(x) {
  m1<-coxph(Surv(tempo3m, evento3m) ~ maternal_age + education2,
  data=bd,x=T)
  m2<-coxph(Surv(tempo3m, evento3m) ~ maternal_age + education2 + bd[,x],
  data=bd,x=T)
  test <- anova(m1, m2, test="Chisq")
  names(test)
  out <- data.frame(test$Df[2], test[[4]][2])
  names(out) <- c("Df", "Pvalue")
  out
}

demog_cols<-seq(13,15) #columns with demographic variables
names(bd[demog_cols])
demog<-as.data.frame(t(sapply(demog_cols, deviance)))
step1.1<-cbind(names(bd[demog_cols]),demog)
step1.1[order(unlist(step1.1$Pvalue)), ]         #Include country

model<-coxph(Surv(tempo3m, evento3m) ~ maternal_age + education2 + country,
  data=bd,x=T)
drop1(model,scope=~country,test="Chisq") # Test variable exclusion
```

Listing C.1: Variable selection algorithm.

```
################## survival analysis #########################
bd$tempo1<-rep(0,nrow(bd))
bd$evento1<-rep(0,nrow(bd))
bd$tempo2<-rep(0,nrow(bd))
bd$evento2<-rep(0,nrow(bd))
bd$tempo3<-rep(0,nrow(bd))
bd$evento3<-rep(0,nrow(bd))
### at 3M
# never BF - will be excluded
bd$tempo1[bd$bap1==1] <- 0
bd$evento1[bd$bap1==1] <- 3

# BF and has stopped
bd$tempo1[bd$bap1==2] <- bd$bap2_sem * 12/52
bd$evento1[bd$bap1==2] <- 1

# Still BF
bd$tempo1[bd$bap1==3] <- bd$idade_basal * 12/52
bd$tempo1[bd$bap1==3] <- 0

bd$tempo2 <- bd$tempo1
bd$evento2 <- bd$evento1

### update at 6M
bd$tempo2[bd$bap4_fu==1] <- 6
bd$evento2[bd$bap4_fu==1] <- 0

bd$tempo3 <- bd$tempo2
bd$evento3 <- bd$evento2

### update at 18M
bd$tempo3[conseguido_18m=1 & mama_18m=0] <- duracao18m_months
bd$evento3[conseguido_18m=1 & mama_18m=0] <- 1

bd$tempo3[conseguido_18m==1 & mama_18m=1] <- idade_18m * 12/52
bd$evento3[conseguido_18m==1 & mama_18m=1] <- 0

############## logistic regression ##############################
bd$pcess<-rep(0,nrow(bd))
bd$pcess[na.omit(which((bd$tempo3m-bd$bap11)>=-0.5))]<-1     # 142 cases, actual > planned
bd$pcess[na.omit(which(bd$evento3m==1 & bd$tempo3m-bd$bap11< -0.5))]<-0    # 267 cases, actual < planned
bd$pcess[na.omit(which(bd$evento3m==0 & bd$bap11>=6 & bd$mama_6m=="No"))]<-0  # 2 cases, actual duration between 3 and 6M < planned
bd$pcess<-factor(bd$pcess,levels=c(0,1),labels=c("failed","achieved"))
```

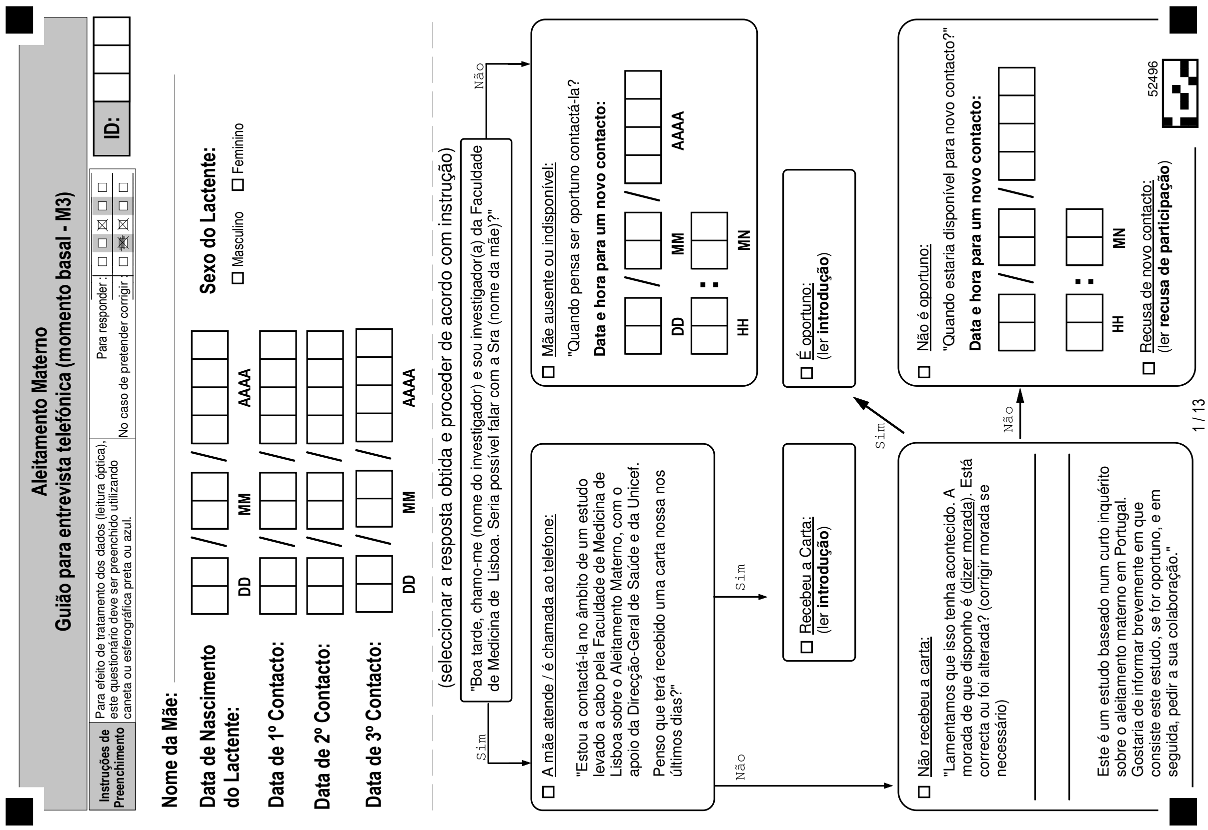
Listing C.2: Response variable construction for BF duration and goal achievement.

```r
# Cox regression
# fitted model - linear term for maternal age
m1<-coxph(Surv(tempo3m, evento3m) ~ maternal_age + education2 + country +
    active_smoker + delivery +cracked_nipples+ mastitis + mother_was_bf +
    hc_used + formulaint + bfondemand+ bses + pacifier1m + nopractices_cat
    + mbrelationship, data=bd,x=T)

# p-spline transformation for maternal age
m2<-coxph(Surv(tempo3m, evento3m) ~ pspline(maternal_age) + education2 +
    country + active_smoker + delivery +cracked_nipples+ mastitis + mother_
    was_bf + hc_used + formulaint + bfondemand+ bses + pacifier1m +
    nopractices_cat + mbrelationship, data=bd,x=T)

# Logistic Regression
# fitted model - linear term for maternal age
m1<- glm(pcess ~ benefits + techniques + maternal_age + education +
    previousbfexp + country + major_complications,family="binomial")
# degree 2 polynomial
m2<-glm(pcess ~ benefits + techniques +  poly(maternal_age, degree=2, raw=
    TRUE)  + education + previousbfexp + country + major_complications,
    family="binomial")
# degree 3 polynomial
m3<-glm(pcess ~ benefits + techniques +  poly(maternal_age, degree=3, raw=
    TRUE)  + education + previousbfexp + country + major_complications,
    family="binomial")
# cubic spline transformation for maternal age
library(splines)
m4 = glm(pcess ~ benefits + techniques +  ns(maternal_age, df=3)  +
    education + previousbfexp + country + major_complications,family="
    binomial")
```

Listing C.3: Fitted models.

Appendix D

Questionnaire



Aleitamento Materno
Guião para entrevista telefónica (momento basal - M3)

Instruções de Preenchimento	Para efeito de tratamento dos dados (leitura óptica), este questionário deve ser preenchido utilizando caneta ou esferográfica preta ou azul.	Para responder: ☐ ☐ ☒ ☐ ☐ No caso de pretender corrigir: ☐ ☒ ☒ ☐ ☐	ID:

Nome da Mãe: ______

Data de Nascimento do Lactente: ☐☐ / ☐☐ / ☐☐☐☐ (DD MM AAAA)

Sexo do Lactente: ☐ Masculino ☐ Feminino

Data de 1º Contacto: ☐☐ / ☐☐ / ☐☐☐☐

Data de 2º Contacto: ☐☐ / ☐☐ / ☐☐☐☐

Data de 3º Contacto: ☐☐ / ☐☐ / ☐☐☐☐ (DD MM AAAA)

(seleccionar a resposta obtida e proceder de acordo com instrução)

"Boa tarde, chamo-me (nome do investigador) e sou investigador(a) da Faculdade de Medicina de Lisboa. Seria possível falar com a Sra (nome da mãe)?"

Sim

☐ A mãe atende / é chamada ao telefone:

"Estou a contactá-la no âmbito de um estudo levado a cabo pela Faculdade de Medicina de Lisboa sobre o Aleitamento Materno, com o apoio da Direcção-Geral de Saúde e da Unicef.

Penso que terá recebido uma carta nossa nos últimos dias?"

Sim → ☐ Recebeu a Carta: (ler **introdução**)

Não → ☐ Não recebeu a carta:

"Lamentamos que isso tenha acontecido. A morada de que disponho é (dizer morada). Está correcta ou foi alterada? (corrigir morada se necessário)

______

______

Este é um estudo baseado num curto inquérito sobre o aleitamento materno em Portugal. Gostaria de informar brevemente em que consiste este estudo, se for oportuno, e em seguida, pedir a sua colaboração."

Sim → ☐ É oportuno: (ler **introdução**)

Não → ☐ Não é oportuno:

"Quando estaria disponível para novo contacto?"

Data e hora para um novo contacto: ☐☐ / ☐☐ / ☐☐☐☐ ☐☐ : ☐☐ (HH MN)

☐ Recusa de novo contacto: (ler **recusa de participação**)

Não

☐ Mãe ausente ou indisponível:

"Quando pensa ser oportuno contactá-la?

Data e hora para um novo contacto: ☐☐ / ☐☐ / ☐☐☐☐ (DD MM AAAA) ☐☐ : ☐☐ (HH MN)

52496

1 / 13

Introdução e consentimento para participação no estudo

"Estamos a realizar um estudo sobre a situação do Aleitamento Materno em Portugal para o qual precisamos de entrevistar mais de 500 mães que tenham dado à luz há cerca de 3 meses. O seu contacto, tal como os contactos das restantes participantes, foi-nos disponibilizado pelo Centro de Genética Médica Dr. Jacinto Magalhães, o Instituto responsável pelo Rastreio Nacional de Doenças Metabólicas, com a sua autorização. Relembro que o seu nome foi seleccionado absolutamente ao acaso para participar neste inquérito com recolha de dados por entrevista telefónica.

Nesta entrevista telefónica serão colocadas questões relacionadas com o nascimento, o aleitamento materno e a família do seu bebé.

Esta entrevista terá a duração aproximada de 15 minutos. E daqui a 3 meses iremos contactá-la novamente para responder a uma segunda entrevista telefónica com a duração de 5 a 10 minutos.

Todos os dados por si fornecidos vão ser tratados de forma confidencial e anónima, não sendo possível identificá-la (nem a si nem ao seu bebé). Os resultados deste estudo serão depois apresentados como estatísticas de especial interesse para médicos e pessoas que trabalhem com crianças.

A sua participação é muito importante para nós, e se decidir participar é livre de mudar de ideias a qualquer momento.

Gostaria de saber se tem qualquer dúvida ou questão sobre este estudo?

Gostaria de saber se aceita participar no estudo respondendo a esta entrevista telefónica?"

Sim

☐ Aceita Participação:

Iniciar **Entrevista Telefónica**

Não

☐ Recusa Participação:

"Compreendo a sua posição e respeito-a inteiramente, por isso não lhe quero roubar mais tempo. Permite-me perguntar-lhe porque decide não participar, de modo a melhoramos o nosso estudo?"

____

____

____

(sugestões se não existir resposta: não tem tempo, pelo tema do estudo, não quer participar...)?

"Peço desculpa pelo incómodo e agradeço o tempo dispendido."

52496

Bloco Elegibilidade

"Vamos começar por lhe colocar algumas questões iniciais que nos garantem que preenche os requisitos para fazer parte do nosso estudo."

1. Após o nascimento, o seu bebé esteve internado numa unidade de cuidados intensivos ou especiais?

☐ Sim (UCI/UCE) -> Agradecer e concluir a entrevista

☐ Não

Infelizmente alguns bebés nascem com problemas ou com malformações

2. O seu bebé nasceu com alguma malformação grave?

☐ Sim -> Agradecer e concluir a entrevista ☐ Não

3. O médico que a acompanha/acompanhou deu-lhe alguma indicação para não iniciar o aleitamento materno?

☐ Sim -> Agradecer e concluir a entrevista ☐ Não

Se **sim**, pode indicar qual foi a razão? __________

Muito obrigada pelas suas respostas. Iremos prosseguir então com a realização da nossa entrevista com a sua permissão.

Bloco Aleitamento Materno

1. Em primeiro lugar, gostaria de lhe perguntar se amamentou ou está a amamentar o seu bebé?

☐ Nunca amamentou -> Passar à pergunta 13

☐ Amamentou e já parou

☐ Ainda está a amamentar

2. Até agora, quantas semanas amamentou no total?

☐☐ Semanas ***Se responder em meses traduzir para o número de semanas***

3. Neste momento, o seu bebé toma algum dos seguintes alimentos ou bebidas regularmente ou de forma continuada (Não contabilizar o consumo episódico/pontual)?

	Sim	Não	NS / NR	Se sim, o bebé quando lhe começou a dar ... Que idade tinha? Até:
a. Água	☐	☐	☐	☐☐ Semanas
b. Chá	☐	☐	☐	☐☐ Semanas
c. Leite artificial (de lata/adaptado)	☐	☐	☐	☐☐ Semanas
d. Outros alimentos (ex. papas, sopa, fruta)	☐	☐	☐	☐☐ Semanas

Se responder em meses traduzir para o número de semanas

Só para quem respondeu que sim ao (c.) leite artificial: (Se não passar à pergunta 5)

4. Existem vários motivos apontados para o início do leite artificial. Dos seguintes, pode dizer-me qual ou quais foram os motivos para a senhora iniciar o leite artificial?
(Mais do que uma opção de resposta)

☐ Sensação de que o bebé ficava com fome

☐ Incerteza sobre a qualidade do seu próprio leite ("leite fraco")

☐ Incerteza sobre a quantidade do seu próprio leite ("leite secou")

☐ Dificuldades na adaptação do bebé à mama ("má pega")

☐ Razões de comodidade

☐ Razões estéticas

☐ Recomendação de profissional de saúde

☐ Sugestão de membro da família

☐ Dificuldade em conciliar com a vida profissional

☐ Complicações mamárias. Qual foi a principal? __________

☐ Aparecimento de uma contra-indicação ao aleitamento materno

☐ Considerou que era altura de começar

☐ Outro motivo. Qual? __________

☐ Não sabe/ não responde

5. O bebé tomou outras vitaminas para além da vitamina D?

☐ Sim. Em que meses? ☐ 1º Mês ☐ 2º Mês ☐ 3º Mês

☐ Não

Em relação à decisão de aleitamento materno....

6. Poderia indicar-me quando tomou a decisão de amamentar? Se foi:

☐ Antes do parto

☐ Após o parto

☐ Não sabe / Não responde

7. E qual ou quais os motivos porque decidiu amamentar?
(Mais do que uma opção de resposta)

a. ☐ Por reconhecimento de vantagens para o bebé

b. ☐ Por reconhecimento de vantagens pessoais maternas

c. ☐ Por recomendação/ conselho médico

d. ☐ Por recomendação/ conselho de outros profissionais de saúde

e. ☐ Por recomendação/ conselho da família

f. ☐ Outro motivo. Qual? __________

g. ☐ Não sabe/ não responde

7.1 (Se mais do que um referido): **Dos que me indicou, qual o motivo principal porque decidiu amamentar?**

☐ Colocar a <u>Letra</u> correspondente

52496

8. Considera que existem vantagens associadas ao aleitamento materno para o bébé?

☐ Sim
☐ Não
☐ Não sabe / Não responde

8.1. Se sim, poderia indicar-me qual ou quais?
(Não ler, solicitar as opções referidas, até a um máximo de 3 opções)

☐ Nutricionais (promoção do crescimento/desenvolvimento)
☐ Imunológicas: anti-infecciosas ou imuno-alérgicas (defesas do organismo)
☐ Psicológicas (reforço auto-estima, sensação de prazer no aleitamento materno)
☐ Prevenção da Diabetes Mellitus
☐ Prevenção da obesidade
☐ Prevenção de linfomas
☐ Prevenção de doenças inflamatórias intestinais crónicas tais como doença de crohn, colite ulcerosa
☐ Reforço da relação mãe-bébé
☐ Outro motivo. Qual? ____________
☐ Não sabe/ não responde

9. E considera que o aleitamento materno lhe traz vantagens pessoalmente?

☐ Sim
☐ Não
☐ Não sabe / Não responde

9.1. Se sim, qual ou quais?
(Não ler, solicitar as opções referidas, até a um máximo de 3 opções)

☐ Facilitação da recuperação física pós parto, incluindo aceleração do retorno ao peso ideal
☐ Diminuição do risco de cancro da mama
☐ Diminuição do risco de cancro do ovário
☐ Diminuição do risco de cancro do útero
☐ Contracepção
☐ Psicológicas (reforço auto-estima, sensação de prazer no aleitamento materno)
☐ Reforço da relação mãe-bébé
☐ Económicas
☐ Outro motivo. Qual? ____________
☐ Não sabe/ não responde

10. Quando decidiu amamentar, até quando pensou fazê-lo?

☐ Até o leite acabar
☐ Até ao bebé querer
☐ Tempo recomendado pelo médico
☐ Até regressar ao trabalho
☐ Outro tempo. Especificar: ____________
☐ Não planeou tempo específico para amamentar
☐ Não sabe / Não responde

11. Para si, isso corresponderia a quantos meses?

Meses

Se não souber ou não responder colocar 99 como resposta

12. Como classificaria a sua motivação para amamentar numa escala de 1 a 5, em que 1 é nenhuma motivação e 5 excelente motivação

Antes do parto?

	☐	☐	☐	☐	☐	
Nenhuma Motivação	1	2	3	4	5	Excelente motivação

Imediatamente após o parto (1a semana)?

	☐	☐	☐	☐	☐	
Nenhuma Motivação	1	2	3	4	5	Excelente motivação

Actualmente, aos 3 meses de vida do bebé?

	☐	☐	☐	☐	☐	
Nenhuma Motivação	1	2	3	4	5	Excelente motivação

Se interrompeu o aleitamento materno, passar à pergunta 14.

Se ainda está a amamentar, passar à pergunta 15.

Apenas para quem nunca amamentou

13. Quando decidiu não amamentar?

☐ Antes do parto
☐ Após o parto
☐ Não sabe / Não responde

52496

14. Dos seguintes motivos frequentemente referidos pelas mulheres, qual ou quais os motivos porque decidiu não amamentar ou interromper o aleitamento materno?
(Mais do que uma opção de resposta)

a. ☐ Sensação de que o bebé ficava com fome
b. ☐ Incerteza sobre a qualidade do seu próprio leite ("leite fraco")
c. ☐ Incerteza sobre a quantidade do seu próprio leite ("leite secou")
d. ☐ Dificuldades na adaptação do bebé à mama ("má pega")
e. ☐ Razões de comodidade
f. ☐ Razões estéticas
g. ☐ Recomendação de profissional de saúde
h. ☐ Sugestão de membro da família
i. ☐ Dificuldade em conciliar com a vida profissional
j. ☐ Complicações mamárias. Qual foi a principal? ____________
l. ☐ Aparecimento de uma contra-indicação ao aleitamento materno
m. ☐ Considerou que já era tempo suficiente
n. ☐ Outro motivo. Qual? ____________
o. ☐ Não sabe/ não responde

14.1 (Se mais do que um referido): **Dos que me indicou, qual o motivo principal porque decidiu não amamentar ou interromper o aleitamento materno?**

Colocar a <u>Letra</u> correspondente

15. Foi aconselhada por alguém ou pela leitura de informação específica a amamentar?

☐ Sim ☐ Não

Se **sim**, por quem ou como (a mãe pode escolher mais que uma opção) - (Dar opções):

☐ Médico ☐ Folhetos de puericultura
☐ Enfermeiro ☐ Revistas/Livros
☐ Aulas de preparação de parto ☐ Não sabe / Não responde
☐ Familiar/Amigos

16. Gostava agora de lhe perguntar qual foi a opinião do pai do bebé em relação ao aleitamento materno?

☐ Favorável
☐ Contra
☐ Indiferente
☐ Não sabe / Não responde

17. E qual foi a opinião da sua mãe em relação ao aleitamento materno?

☐ Favorável
☐ Contra
☐ Indiferente
☐ Não sabe / Não responde

18. A senhora quando nasceu foi amamentada?

☐ Sim ☐ Não ☐ Não sabe / Não responde

Bloco Gravidez

Agora iremos colocar-lhe algumas perguntas relativamente aos seus filhos e à sua gravidez…

1. Quantos filhos tem, incluindo este?

Filhos

Se é o primeiro filho ir para pergunta 4

2. Com que idade teve o primeiro filho?

Anos

3. Dos seus outros filhos, deu de mamar a quantos?

Filhos

3.1 Se **pelo menos um**, qual o tempo médio de aleitamento materno:

Meses

3.2 E, de uma maneira geral, sentiu-se satisfeita com essa(s) experiência(s)?

☐ Sim ☐ Não ☐ Não sabe / Não responde

Em relação a esta última gravidez…

4. Foi planeada?

☐ Sim ☐ Não ☐ Não sabe / Não responde

5. Quantas consultas médicas teve durante esta última gravidez?

Consultas

52496

Se não teve nenhuma consulta saltar para a pergunta 7

6. Com quantas semanas de gravidez estava quando foi à primeira consulta?

☐☐ Semanas

6.1 Onde se realizaram a maioria das consultas?

☐ Centro de saúde ☐ Maternidade/Hospital público
☐ Hospital privado ☐ Consultório particular / Clínica Privada
☐ Outro local. Qual? ______
☐ Não sabe/ não responde

7. Frequentou algum tipo de curso de preparação para o parto?

☐ Sim ☐ Não ☐ Não sabe / Não responde

7.1 Se sim, por quem foram dadas essas aulas?

☐ Médico
☐ Enfermeiro
☐ Outro. Quem? ______
☐ Não sabe / Não responde

7.2 Se sim, como classifica numa escala de 1 a 5 a sua satisfação com o apoio prestado? (1 muito má a 5 muito boa)

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

8. Gostaria agora de lhe perguntar se, <u>durante a gravidez</u>, esteve exposta ao tabaco?

☐ Sim ☐ Não ☐ Não sabe / Não responde

8.1 Se sim, por conviver com fumadores ou por fumar?

☐ Fumadora Activa
☐ Fumadora Passiva
☐ Não sabe / Não responde

(Se respondeu **Fumadora Activa**), em média quantos cigarros fuma por dia?

☐☐ Cigarros

Bloco Parto

1. Onde ocorreu o parto? Especificar onde.

☐ Hospital Público. Qual? ______
☐ Hospital Particular. Qual? ______
☐ Em casa
☐ Outro sítio. Qual? ______
☐ Não sabe / Não responde

2. Que tipo de parto teve?

☐ Normal, vaginal
☐ Cesariana -> Passar à pergunta 6
☐ Outro. Qual? ______
☐ Não sabe / Não responde

3. O seu trabalho de parto foi provocado?

☐ Sim ☐ Não ☐ Não sabe / Não responde

4. Durante o parto, foi administrada epidural?

☐ Sim ☐ Não ☐ Não sabe / Não responde

As próximas questões dizem respeito ao período logo após o parto.

5. Durante quanto tempo esteve internada na maternidade após o parto?

☐ < 24 horas
☐ 24-48 horas
☐ > 48 horas
☐ Não teve parto no hospital -> Passar ao bloco seguinte
☐ Não sabe / Não responde

6. Teve algum problema de saúde após o parto?

☐ Sim ☐ Não ☐ Não sabe / Não responde

7. O seu bebé foi colocado junto de si logo na primeira hora de vida?

☐ Sim ☐ Não ☐ Não sabe / Não responde

52496

6 / 13

8. Quanto tempo passou entre o parto e a 1ª mamada?

☐☐☐ Minutos ***Se responder em horas traduzir para o número de minutos***

9. Durante a estadia na maternidade, o seu bebé esteve a maior parte do tempo junto de si?

☐ Sim ☐ Não ☐ Não sabe / Não responde

10. Na maternidade, o bebé mamava sempre que queria?

☐ Sim ☐ Não ☐ Não sabe / Não responde

11. Na maternidade, que alimentos ou bebidas foram dados ao seu bebé? (ler opções)

	Sim	Não	NS / NR
a. Leite da mãe	☐	☐	☐
b. Leite artificial	☐	☐	☐
c. Água	☐	☐	☐
d. Chá	☐	☐	☐
e. Outros. Qual?	☐	☐	☐

______

12. Na maternidade, foram-lhe oferecidos kits de amostras ou publicidade a marcas de leite artificial?

☐ Sim ☐ Não ☐ Não sabe / Não responde

13. Na maternidade, algum profissional de saúde lhe ensinou como amamentar?

☐ Sim ☐ Não ☐ Não sabe / Não responde

14. Na maternidade, algum profissional de saúde lhe explicou como manter o aleitamento se precisasse de estar separada do seu filho?

☐ Sim ☐ Não ☐ Não sabe / Não responde

Bloco Acompanhento de Aleitamento Materno

1. Alguma vez, algum profissional de saúde lhe falou sobre os benefícios do aleitamento materno?

☐ Sim ☐ Não ☐ Não sabe / Não responde

1.1 Se **sim**, quando?
(pode assinalar mais do que uma opção - **ler as opções**)

☐ <u>Nas consultas durante a gravidez. Quem?</u>

☐ Médico de família ☐ Médico obstetra ☐ Enfermeiro
☐ Outro. Quem? ______
☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

☐ <u>Nas aulas de preparação para o parto. Quem?</u>

☐ Médico de família ☐ Médico obstetra ☐ Enfermeiro
☐ Outro. Quem? ______
☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

☐ <u>Na maternidade. Quem?</u>

☐ Médico obstetra ☐ Pediatra ☐ Enfermeiro
☐ Outro. Quem? ______
☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

52496

7 / 13

☐ Nas consultas após o parto. Quem?

☐ Médico de família ☐ Pediatra ☐ Médico obstetra

☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

2. Algum profissional de saúde lhe falou sobre as dificuldades no aleitamento materno, tais como: complicações mamárias, técnica correcta para amamentar ou outras e como superar essas dificuldades?

☐ Sim ☐ Não ☐ Não sabe / Não responde

2.1 Se **sim**, quando? (pode assinalar mais do que uma opção) - **ler as opções**

☐ Nas consultas durante a gravidez. Quem?

☐ Médico de família ☐ Médico obstetra ☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não sabe / Não responde

☐ Nas aulas de preparação para o parto. Quem?

☐ Médico de família ☐ Médico obstetra ☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

☐ Na maternidade. Quem?

☐ Pediatra ☐ Médico obstetra ☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

☐ Nas consultas após o parto. Quem?

☐ Médico de família ☐ Médico obstetra ☐ Pediatra

☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

E como classifica numa escala de 1 a 5 a sua satisfação relativamente a essa informação?

Muito Má ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Boa

☐ Não Sabe / Não Responde

3. Tem médico de família?

☐ Sim ☐ Não ☐ Não sabe / Não responde

4. Está inscrita em algum Centro de Saúde?

☐ Sim

☐ Não -> Passar à pergunta 6

☐ Não sabe / Não responde -> Passar à pergunta 6

5. No seu centro de saúde existiam serviços de apoio ou espaços próprios para o aleitamento materno (espaço amamentação)?

☐ Sim ☐ Não ☐ Não sabe / Não responde

5.1. Recorreu a algum desses espaços ou serviços de apoio ao Aleitamento Materno?

☐ Sim ☐ Não ☐ Não sabe / Não responde

5.1.1 Se sim, quando? (pode seleccionar mais que uma opção)

☐ Na gravidez ☐ Depois do parto ☐ NS / NR

5.1.2 Em que consistiam esses serviços/espaço de apoio? (a mãe pode escolher mais que uma opção)

☐ Sessões individuais

☐ Sessões em grupo

☐ Espaço privado para poder dar de mamar -> Se apenas esta passar à pergunta 6

☐ Outro. O quê? ____________________

5.1.3 Que assuntos abordavam? (pode seleccionar mais que uma opção)

☐ Benefícios do Aleitamento Materno

☐ Complicações mamárias que podem surgir durante o AM

☐ Importância do Aleitamento Materno exclusivo

☐ Resolução de dificuldades ou problemas que podem surgir durante o Aleitamento Materno

☐ Treino do posicionamento do bebé e pega correcta

☐ Como manter o AM caso estivesse separada do seu bebé

☐ Esclarecimento de dúvidas

☐ Outra. Qual? ____________________

5.1.4 Quem prestou apoio nas sessões?(pode escolher mais que uma opção)

☐ Médico de família

☐ Enfermeiro

☐ Outro. Quem? ____________________

☐ Não sabe / Não responde

5.1.5 Como classifica numa escala de 1 a 5 a sua satisfação com sessões em geral? (1 muito má a 5 muito boa)

Muito Mau ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Bom

☐ Não Sabe / Não Responde

6. Está informada acerca da existência de grupos de apoio ao AM, fora do seu centro de saúde?

☐ Sim ☐ Não ☐ Não sabe / Não responde

6.1 Como tomou conhecimento da existência desses grupos? (pode escolher mais que uma opção)

☐ Consultas da gravidez ☐ Internet

☐ Maternidade ☐ Amigos

☐ Consultas depois do parto ☐ Publicidade/panfletos

☐ Outro meio. Qual? ____________________

☐ Não sabe / Não responde

6.2 Recorreu a algum serviço desses grupos?

☐ Sim ☐ Não ☐ Não sabe / Não responde

6.2.1 Como classifica numa escala de 1 a 5 a sua satisfação com apoio prestado? (1 muito má a 5 muito boa)

Muito Mau ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 Muito Bom

☐ Não Sabe / Não Responde

Bloco Conhecimento do Aleitamento Materno

Em relação ao período de aleitamento materno....

(se a mãe nunca amamentou saltar este bloco)

1. Enquanto amamenta/amamentava, que técnicas ou truques pratica/costumava praticar para facilitar a amamentação? (Não ler, solicitar as opções referidas, até a um máximo de 3 opções)

☐ Procurar um ambiente calmo para amamentar

☐ Colocar-se a si própria numa posição confortável

☐ Segurar o bebé junto a si, suportando o seu dorso, ombros e pescoço

☐ Alinhar a cabeça e o corpo do bebé

☐ Assegurar que o corpo do seu bebé está encostado ao seu, barriga com barriga

☐ Garantir que o bebé chega facilmente à mama, sem ter de girar a cabeça ou pescoço

☐ Garantir que a boca do bebé está bem aberta

☐ Alinhar o nariz do bebé com o mamilo

☐ Garantir que o bebé mama na aréola e não apenas no mamilo

☐ Outra técnica. Qual? ____________________

2. Dos comportamentos que lhe vou ler, podia-me indicar quais pratica (ou praticava durante o período em que amamenta(va)?

- ☐ Extracção mamária de leite?
  - Se **sim**:
  - ☐ Bomba manual ☐ Bomba eléctrica ☐ Mamilos artificiais
  - ☐ Outros. Especifique: ______
  - ☐ Não sabe / Não responde
- ☐ Uso de um biberão?
  - Se **sim,** a partir de que idade? ☐☐ Semanas
- ☐ Uso de chupeta?
  - Se **sim**: Quando colocou a chupeta ao bebé pela primeira vez?
  - ☐ Na maternidade
  - ☐ Após a alta da maternidade. Quando? ☐☐ Semanas
  - ☐ Não sabe / Não responde
- ☐ Pedir ajuda no que respeita ao aleitamento materno?
  - Se **sim,** a quem? (pode escolher mais que uma opção)
  - ☐ Médico de família
  - ☐ Médico obstetra
  - ☐ Pediatra
  - ☐ Enfermeiro
  - ☐ À sua mãe
  - ☐ Amigas
  - ☐ Grupo apoio amamentação
  - ☐ Outro: quem? ______
- ☐ Não sabe / Não responde

3. E se alguma vez tivesse um problema relacionado com o aleitamento materno era fácil pedir ajuda ao seu médico ou outro profissional de saúde?

☐ Sim ☐ Não ☐ Não sabe / Não responde

Se **sim,** a quem recorreria? (pode escolher mais que uma opção)

- ☐ Médico de família
- ☐ Médico obstetra
- ☐ Pediatra
- ☐ Enfermeiro
- ☐ Outro: quem? ______
- ☐ Não sabe / Não responde

4. Durante o período de aleitamento materno tem/teve alguém que a auxiliasse na execução das tarefas domésticas (tais como: lavar a loiça, aspirar a casa, passar a ferro, etc..)

☐ Sim ☐ Não ☐ Não sabe / Não responde

5. Conhece algumas complicações mamárias que possam surgir durante o aleitamento materno?

☐ Sim ☐ Não ☐ Não sabe / Não responde

5.1 Se **sim**, quais? (Não ler, solicitar as opções referidas, até a um máximo de 3 opções)

- ☐ Mamas muito cheias e dolorosas (correspondente a ingurgitamento mamário)
- ☐ Inchaços mamários localizados, tipo caroços, dolorosos (correspondentes a bloqueio do ductos)
- ☐ Infecção mamária (mastite)
- ☐ Fissuras mamárias
- ☐ Mamilos planos ou invertidos
- ☐ Produção de leite em quantidade insuficiente
- ☐ Outras. Quais? ______
- ☐ Não sabe / Não responde

6. Teve alguma complicação mamária neste período de aleitamento materno?

☐ Sim ☐ Não ☐ Não sabe / Não responde

6.1 Se **sim**, quais? (Não ler, assinalar as opções referidas, até a um máximo de 3 opções")

- ☐ Mamas muito cheias e dolorosas (correspondente a ingurgitamento mamário)
- ☐ Inchaços mamários localizados, tipo caroços, dolorosos (correspondentes a bloqueio do ductos)
- ☐ Infecção mamária (mastite)
- ☐ Fissuras mamárias
- ☐ Mamilos planos ou invertidos
- ☐ Produção de leite em quantidade insuficiente
- ☐ Outras. Quais? ______
- ☐ Não sabe / Não responde

52496

10 / 13

Bloco Escala Auto-Eficácia

Vou ler-lhe algumas afirmações que remetem para o grau de confiança relativamente à amamentação do seu bebé e pedir-lhe para me indicar, numa escala de 1 a 5, em que 1 é nada confiante e 5 é muito confiante, qual o número que melhor descreve como se sente

1 = nada confiante 2 = pouco confiante 3 = às vezes confiante 4 = confiante 5 = muito confiante

	1	2	3	4	5
1. Eu consigo sempre perceber se o meu bebé está a receber leite suficiente	☐	☐	☐	☐	☐
2. Eu consigo sempre lidar com sucesso com a amamentação, tal como com outras tarefas desafiantes	☐	☐	☐	☐	☐
3. Eu consigo sempre amamentar o meu bebé sem ser necessário dar-lhe suplemento	☐	☐	☐	☐	☐
4. Eu consigo sempre garantir que o meu bebé faz uma boa pega durante toda a mamada	☐	☐	☐	☐	☐
5. Eu consigo sempre gerir de forma satisfatória a amamentação	☐	☐	☐	☐	☐
6. Eu consigo sempre gerir a amamentação mesmo que o meu bebé esteja a chorar	☐	☐	☐	☐	☐
7. Eu posso sempre continuar a desejar amamentar	☐	☐	☐	☐	☐
8. Eu consigo sempre sentir-me à vontade a amamentar o meu bebé quando estão presentes outros membros da família	☐	☐	☐	☐	☐
9. Eu consigo sempre sentir-me satisfeita com a minha experiência de amamentar	☐	☐	☐	☐	☐
10. Eu consigo sempre lidar bem com o facto de a amamentação poder consumir tempo	☐	☐	☐	☐	☐
11. Eu consigo sempre amamentar o meu bebé esvaziando uma mama antes de passar à outra	☐	☐	☐	☐	☐
12. Eu consigo sempre continuar a amamentar o meu bebé em todas as mamadas	☐	☐	☐	☐	☐
13. Eu consigo sempre amamentar o meu bebé quando ele pede	☐	☐	☐	☐	☐
14. Eu consigo sempre identificar quando é que o meu bebé acabou de mamar	☐	☐	☐	☐	☐

52496

11 / 13

Bloco Dados Sócio-Demográficos

Para terminar gostaria de lhe pedir mais apenas mais algumas informações de carácter mais geral:

1. Qual é o seu estado civil?

☐ Solteira
☐ Casada ou em união de facto (vive maritalmente há pelo menos 2 anos)
☐ Divorciada ou separada
☐ Viúva
☐ Não sabe / Não responde

2. Em que país nasceu?

☐ Portugal
☐ Brasil
☐ País Africano de Língua Oficial Portuguesa (PALOP)
☐ Europa Ocidental
☐ Europa do Leste
☐ Outros países da América do Sul e Central
☐ América do Norte
☐ Outros países de África
☐ Ásia
☐ Não sabe / Não responde

2.1 Se nasceu em Portugal, por favor indique em que distrito:

☐ Aveiro ☐ Portalegre
☐ Beja ☐ Porto
☐ Braga ☐ Santarém
☐ Bragança ☐ Setúbal
☐ Castelo Branco ☐ Viana do Castelo
☐ Coimbra ☐ Vila Real
☐ Évora ☐ Viseu
☐ Faro ☐ Região Autónoma da Madeira
☐ Guarda ☐ Região Autónoma dos Açores
☐ Leiria ☐ Não sabe / Não responde
☐ Lisboa

2.2. Se não nasceu em Portugal, há quantos anos reside em Portugal?

☐☐ Anos

3. Como é que descreveria o seu grupo étnico/raça?

☐ Branco
☐ Negro ou Mulato
☐ Asiático
☐ Latino-americano
☐ Outro (Qual?) ______ ☐☐
☐ Não sabe / Não responde

4. Qual o nível de ensino mais elevado que frequenta ou frequentou?

☐ Nenhum (0 anos)
☐ 1º Ciclo do ensino básico 4ª classe (1-4 anos)
☐ 2º Ciclo do ensino básico Preparatório (5-6 anos)
☐ 3º Ciclo do ensino básico 5º ano dos liceus (7-9 anos)
☐ Ensino secundário 7º ano dos liceus (10-12 anos)
☐ Ensino pós-secundário não superior (ou não terciário)
☐ Ensino superior universitário (bacharelato, licenciatura, mestrado)
☐ Doutoramento
☐ Não sabe / Não responde

5. E quantos anos de escolaridade completou com aproveitamento?

☐☐ Anos

Se não souber ou não responder colocar 99 anos como resposta

52496

6. Das seguintes categorias, qual a que melhor descreve a sua ocupação principal actual?

☐ Exerce uma profissão
Qual? ______ ☐☐
☐ Estudante -> Passar à pergunta 10
☐ Ocupa-se das tarefas domésticas -> Passar à pergunta 10
☐ À procura do primeiro emprego -> Passar à pergunta 10
☐ Desempregada
☐ Reformada
☐ Permanentemente incapacitada
☐ Outra situação
☐ Não sabe / Não responde -> Passar à pergunta 10

7. E qual é (era) a situação nessa profissão?

☐ Trabalha por conta de outrem
☐ Trabalha por conta própria
☐ Trabalha por conta própria como empregador(a)
☐ Trabalha para uma pessoa de família sem receber
☐ Outra situação
☐ Não sabe / Não responde

8. Diga-me, por favor, o número total de horas semanais que trabalha(va)?

☐☐ Horas

Se não souber ou não responder colocar 99 anos como resposta

9. Contando consigo, quantas pessoas - incluindo o recém-nascido e outras crianças- vivem habitualmente em sua casa?

☐☐ Pessoas

Se não souber ou não responder colocar 99 anos como resposta

[Se mora com alguém]

10. Podia dizer-me todas as pessoas com quem vive (a mãe do bebé)? (incluindo o recém-nascido)

☐ Pai do bébé
☐ Outro cônjuge/companheiro
☐ Mãe
☐ Pai
☐ Sogro/sogra
☐ Avô/Avó
☐ Filho/Filha
☐ Irmão/Irmã
☐ Outra pessoa de família
☐ Outra pessoa

Chegámos ao fim desta primeira fase do estudo, esperamos que tenha gostado de colaborar connosco!

Muito obrigado pela sua participação.

Voltaremos a contactá-la daqui a 3 meses para responder ao segundo questionário que terá a duração de apenas 5 a 10 minutos .

No entanto gostariamos de saber se:

Há alguma altura do dia que prefira ser contactada?
Especificar:

☐☐ : ☐☐
HH **MM**

Prefere que liguemos para este número de telefone ou outro?
Especificar:

☐ Mesmo número
☐ Outro. Qual?

☐☐☐☐☐☐☐☐☐

52496

References

Al-Sahab, B., Lanes, A., Feldman, M. & Tamim, H. (2010). Prevalence and predictors of 6-month exclusive breastfeeding among Canadian women: a national survey. *BMC pediatrics*, **10**, 20. 2, 47

Bai, D.L., Wu, K.M. & Tarrant, M. (2013). Association between intrapartum interventions and breastfeeding duration. *Journal of Midwifery & Women's Health*. 3, 47

Caldeira, A.P., Goulart, E. *et al.* (2000). Breastfeeding in montes claros, minas gerais: a representative sample study. *J Pediatr (Rio J)*, **76**, 65–72. 2

Carvalho, M.S., Andreozzi, V.L., Codeço, C.T., Campos, D.P., Barbosa, M.T.S. & Shimakura, S.E. (2011). *Análise de Sobrevivência - Teoria e aplicações em saúde*. Editora Fiocruz, Rio de Janeiro, 2nd edn. 14, 17, 45, 77

Collett, D. (2003). *Modelling survival data in medical research*, vol. 57. CRC press. 8, 11, 12, 14, 15, 16

Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)*, **34**, 187–220. 11

Declercq, E., Labbok, M.H., Sakala, C. & O'Hara, M. (2009). Hospital practices and women's likelihood of fulfilling their intention to exclusively breastfeed. *Journal Information*, **99**. 47

Dobson, A.J. (2002). *An introduction to generalized linear models*. Chapman & Hall. 50, 51, 52

Eidelman, A.I., Schanler, R.J., Johnston, M., Landers, S., Noble, L., Szucs, K., Viehmann, L. *et al.* (2012). Breastfeeding and the use of human milk. *Pediatrics*, **129**, e827–e841. 1

Forster, D.a., McLachlan, H.L. & Lumley, J. (2006). Factors associated with breastfeeding at six months postpartum in a group of Australian women. *International breastfeeding journal*, **1**, 18. 2, 47

Fox, J. & Weisberg, S. (2010). *An R companion to applied regression*. SAGE Publications, Incorporated. 51, 52

Fuchs, S.C., Victora, C.G. & Fachel, J. (1996). Modelo hierarquizado: uma proposta de modelagem aplicada à investigação de fatores de risco para diarreia grave. *Revista de Saúde Pública*, **30**, 168–178. 19

Greenland, S. (1989). Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, **79**, 340–349. 59

Hanson, L. (2000). The mother-offspring dyad and the immune system. *Acta paediatrica*, **89**, 252–258. 1

Harrell, F.E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer. 8, 14, 63

Hector, D., King, L., Webb, K. & Heywood, P. (2005). Factors affecting breastfeeding practices. applying a conceptual framework. *New South Wales public health bulletin*, **16**, 52–55. 2

Hosmer, D.W. & Lemeshow, S. (2000). *Applied logistic regression*, vol. 354. Wiley-Interscience. 49, 50, 51, 52, 53, 67

Howard, C.R., Howard, F.M., Lanphear, B., Eberly, S., Oakes, D., Lawrence, R.A. *et al.* (2003). Randomized clinical trial of pacifier use and bottle-feeding or cupfeeding and their effect on breastfeeding. *Pediatrics*, **111**, 511–518. 48

Kleinbaum, D.G. & Klein, M. (2005). *Survival analysis, A Self-Learning Text*. Springer, 2nd edn. 11

Kleinbaum, D.G. & Klein, M. (2012). *Survival analysis, A Self-Learning Text*. Springer, 3rd edn. 8

Kramer, M.S., Barr, R.G., Dagenais, S., Yang, H., Jones, P., Ciofani, L. & Jané, F. (2001). Pacifier use, early weaning, and cry/fuss behavior: a randomized controlled trial. *JAMA : the journal of the American Medical Association*, **286**, 322–6. 48

Leite, A. (2012). *Análise longitudinal da mudança terapêutica nos doentes seropositivos para o VIH*. Master's thesis, Faculdade de Ciências da Universidade de Lisboa. 77

Mattar, C.N., Chong, Y.S., Chan, Y.S., Chew, A., Tan, P., Chan, Y.H. & Rauff, M.H.J. (2007). Simple antenatal preparation to improve breastfeeding practice: a randomized controlled trial. *Obstetrics & Gynecology*, **109**, 73–80. 68

Orfão, A. (2012). Registo do Aleitamento Materno | Relatório julho 2010 – junho 2011. Tech. rep., Direção-Geral da Saúde, Divisão de Saúde Reprodutiva; Mama Mater | Associação pró Aleitamento Materno em Portugal. 1

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. 77

Rasenack, R., Schneider, C., Jahnz, E., Schulte-Mönting, J., Prömpeler, H. & Kunze, M. (2012). Factors associated with the duration of breastfeeding in the freiburg birth collective, germany (freistill). *Geburtshilfe und Frauenheilkunde*, **72**, 64–69. 2

Rocha, C. & Papoila, A.L. (2009). *Análise de Sobrevivência*. Sociedade Portuguesa a e Estatística, Lisboa. 8, 11, 12, 46

Scott, J.A., Binns, C.W., Oddy, W.H. & Graham, K.I. (2006). Predictors of breastfeeding duration: evidence from a cohort study. *Pediatrics*, **117**, e646–e655. 3, 47

Skafida, V. (2012). Juggling work and motherhood: the impact of employment and maternity leave on breastfeeding duration: a survival analysis on Growing Up in Scotland data. *Maternal and child health journal*, **16**, 519–27. 3

Su, L.L., Chong, Y.S., Chan, Y.H., Chan, Y.S., Fok, D., Tun, K.T., Ng, F.S. & Rauff, M. (2007). Antenatal education and postnatal support strategies for improving rates of exclusive breast feeding: randomised controlled trial. *bmj*, **335**, 596. 68

Therneau, T.M. (2013). *A Package for Survival Analysis in S*. 12

Therneau, T.M. & Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer. 12, 14

Thulier, D. & Mercer, J. (2009). Variables associated with breastfeeding duration. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, **38**, 259–268. 2, 19

Torvaldsen, S., Roberts, C.L., Simpson, J.M., Thompson, J.F., Ellwood, D.A. *et al.* (2006). Intrapartum epidural analgesia and breastfeeding: a prospective cohort study. *Int Breastfeed J*, **1**, 1–7. 47

Turkman, M.A.A. & Silva, G. (2000). *Modelos Lineares Generalizados – da teoria à prática*. Sociedade Portuguesa a e Estatística, Lisboa. 19

Victora, C.G., Huttly, S.R., Fuchs, S.C. & Olinto, M.T. (1997). The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *International journal of epidemiology*, **26**, 224–7. 19

Village, E.G. (2012). Breastfeeding and the use of human milk. *Pediatrics*, **129**, e827–41. 1, 33

Yngve, A. & Sjöström, M. (2001). Breastfeeding determinants and a suggested framework for action in Europe. *Public Health Nutrition*, **4**. 2